THE UNIVERSITY
of EDINBURGH

An analysis of the sequence features contributing to centromere organisation and CENP-A positioning and incorporation

Nicholas R.T. Toda

Thesis for the Degree of Doctor of Philosophy

University of Edinburgh

2015

This thesis was composed by myself and the research presented is my own unless otherwise stated. This work has not been submitted as part of any previous degree assessment.

Nicholas Toda
2015

Table of contents

Abstract

Centromere identity is integral for proper kinetochore formation and chromosome segregation. In most species chromosomes have a centromere at a defined locus that is propagated across generations. The histone H3 variant CENP-A acts as an epigenetic mark for centromere identity in most species studied. CENP-A is absent from the inactivated centromere on dicentric chromosomes and present at neocentromeres that form on non-centromeric sequences. Thus, the canonical centromere sequence is neither necessary nor sufficient for centromere function. Nevertheless, centromeres are generally associated with particular sequences. Understanding the organisation of centromeric sequence features will provide insight into centromere function and identity.

In this study I use the fission yeast *Schizosaccharomyces pombe* model system to address the relationship between CENP-A^{Cnp1} and centromeric sequence features. These analyses reveal that CENP-A^{Cnp1} nucleosomes are highly positioned within the central domain by large asymmetric AT-rich gaps. The same sequence features underlying CENP-A^{Cnp1} positioning are conserved in the related species *S. octosporus*, but are not found at neocentromeres, suggesting that they are important but non-essential for centromere function. CENP-A^{Cnp1} over-expression leads to ectopic CENP-A^{Cnp1} incorporation primarily at sites associated with heterochromatin, including the sites where stable neocentromeres form. Ectopic CENP-A^{Cnp1} also occupies additional sites within the central domain that are not occupied in cells with wild-type CENP-A^{Cnp1} levels. In wild-type cells CENP-A^{Cnp1} occupied sites are likely also occupied by H3 nucleosomes or the CENP-T/W/S/X nucleosome-like complex in a mixed population. Several candidate proteins were investigated to determine a protein residing in the large gaps between CENP-A^{Cnp1} nucleosomes could be identified. No proteins could be localised to the AT-rich gaps between CENP-A^{Cnp1} nucleosomes, but the origin recognition complex in a promising candidate.

The results presented in this thesis demonstrate that nucleosomes within the fission yeast centromere central domain are highly positioned by sequence features in a conserved manner. This positioning also allows for another complex, possibly the origin recognition complex, to bind to DNA. Nucleosome positioning, DNA replication, and transcription could individually and collectively influence CENP-A^{Cnp1} assembly and centromere function. Further experiments in fission yeast will

continue to provide insight into the general properties of centromere function and identity.

Abbreviations

ARS	autonomously replicating sequence
ATP	adenosine triphosphate
bp	base pair
BLAST	Basic Local Alignment Search Tool
CATD	CENP-A targeting domain
cc	central core
CCAN	Constitutive Centromere Associated Network
cdc	cell division cycle
CDE	centromere DNA element
CENP	centromere protein
ChIP	chromatin immunoprecipitation
ChIP-chip	ChIP-on-chip
ChIP-qPCR	ChIP coupled to detection by quantitative real-time PCR
ChIP-seq	chromatin immunoprecipitation sequencing
Cre	Cre recombinase
C-terminal	carboxy-terminal
dH_20	distilled water
DNA	deoxyribonucleic acid
dsRNA	double-stranded RNA
ENC	evolutionary new centromere
EM	electron microscopy
FRET	Förster resonance energy transfer
GFP	green fluorescent protein
H3K9me2	histone H3 dimethylated on lysine 9
HDAC	histone deacetylase
HJURP	Holliday junction recognition protein
HP1	heterochromatin protein 1
IgG	immunoglobulin
imr	inner-repeat
IP	immunoprecipitation
kb	kilobase
kD	kilodalton
LacI	lac repressor
LacO	lac operon

lox	Lox site
MCM	minichromosome maintenance
MNase	micrococcal nuclease
MT	microtubule
nmt	no message in thiamine
N-terminal	amino-terminal
ORC	Origin Recognition Complex
otr	outer-repeat
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMG	*pombe* media glutamate
qRT-PCR	quantitative real -time PCR
rDNA	ribosomal DNA
RDRC	RNA-directed RNA Polymerase Complex
RdRP	RNA-dependant RNA polymerase
RISC	RNA induced silencing complex
RITS	RNA-mediated Initiation of Transcriptional Silencing
RNA	ribonucleic acid
RNAi	RNA interference
RNAPII	RNA polymerase II
RNAPIII	RNA polymerase III
rpm	revolutions per minute
siRNA	small interfering RNA
tRNA	transfer RNA
YES	Yeast Extract Supplemented

List of Figures and Tables

Chapter 1: Introduction

Kinetochores are assembled at the centromere to mediate microtubule attachments to the mitotic and meiotic spindle and promote proper chromosome segregation during cell division. Accurate centromere identification, kinetochore formation, and microtubule attachment are integral to successful cell division. Errors in any of these processes can lead to aneuploidy, chromosomes breakage, and cell death. Ultimately these errors can lead to genetic abnormalities, disease, and cancer in humans with potential concomitant lethality.

1.1 Centromere identity and organisation

Sequence dependent point centromeres of budding yeasts

Centromere structure varies widely between species with the budding yeast *S. cerevisiae* having sequence dependent centromere formation at a defined locus, termed a point centromere (Figure 1.1; Allshire and Karpen, 2008). Isolation and subcloning of yeast centromere sequences revealed that short sequences could confer centromere function (Clarke and Carbon, 1980; Fitzgerald-Hayes et al., 1982). Nevertheless, centromere sequences did not share large-scale homologies between different centromeres. However, sequence analysis revealed short motifs that had conserved relative arrangement with slightly variable total lengths (~125 bp long) within different centromeres. These motifs were termed CDEs (centromere DNA elements) with CDEI and CDEIII separated by a highly asymmetrically AT-rich element CDEII. The 125 bp sequence conferred functional centromere activity on both plasmids and linear artificial chromosomes (Cottarel et al., 1989). Although not homologous between centromeres at a sequence level, the AT-content and DNA bending of CDEII, as well as the orientation of CDEIII, contribute to centromere functionality and proper chromosome segregation (Murphy et al., 1991). The CDEII sequence itself is not necessary for centromere function and sequences of similar homopolymer composition show a similar level of functionality, while disruption of homopolymer composition disrupted centromere function (Baker and Rogers, 2005). Importantly, CDEI and CDEIII are highly conserved between chromosomes and single and double point mutations in these elements can disrupt centromere activity,

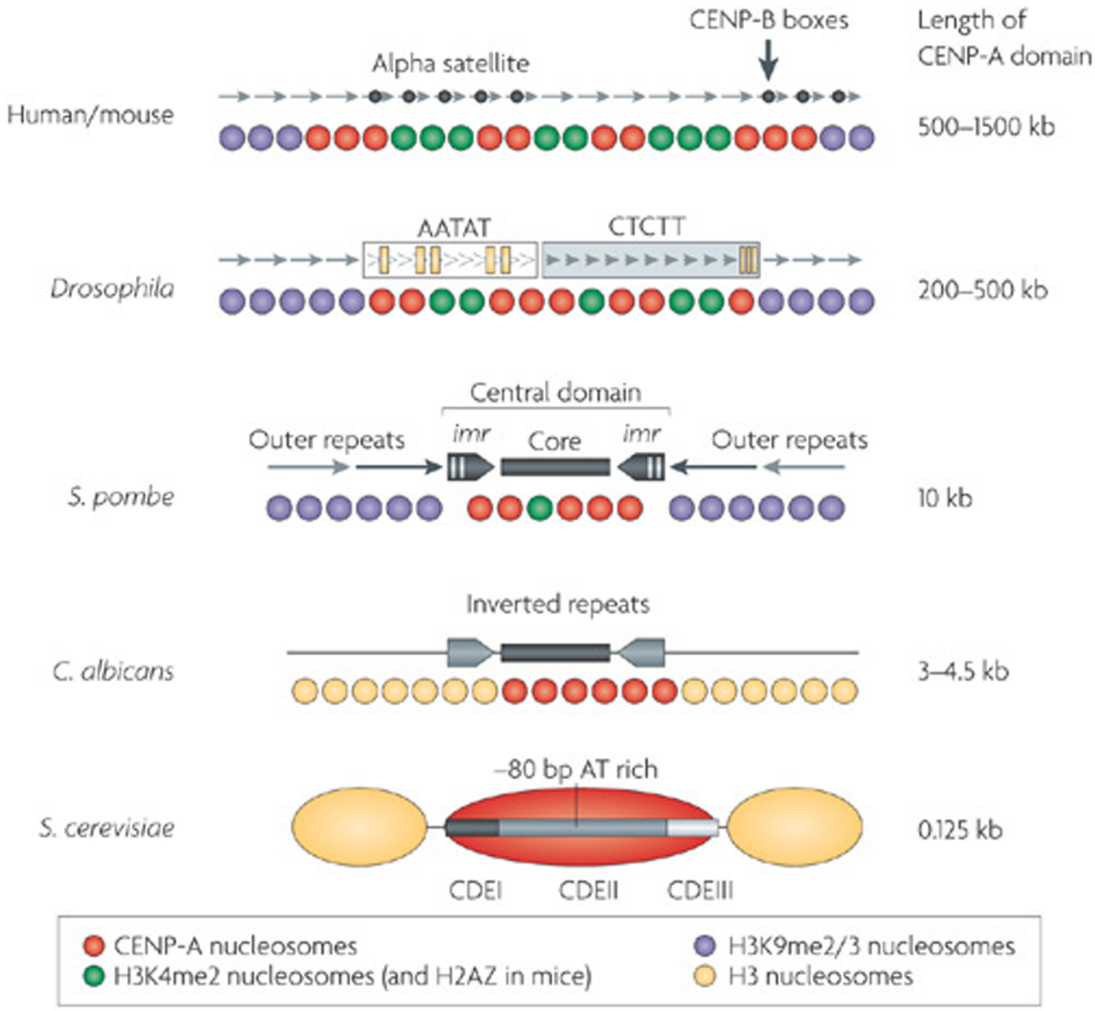


Figure 1.1. Centromere structure varies widely (Allshire and Karpen, 2008). Centromere structure in a number of model species. CENP-A nucleosomes are shown in red. Human and mouse centromeres form on alpha satellite repeat arrays and contain CENP-B boxes to allow CENP-B binding. *Drosophila* centromeres also form on repetitive sequences. Fission yeast *Schizosaccharomyces pombe* centromeres comprise a non-repetitive central core flanked by inverted inner-repeats (imr) and heterochromatic outer-repeats. CENP-A^{Cnp1} occupies the central core and part of the imr, collectively termed the central domain. Budding yeast centromeres are small sequence specific point centromeres that contain a single CENP-A^{Cse4} nucleosome.

indicating that centromere identity is sequence dependent in *S. cerevisiae* (Hegemann et al., 1988).

The discovery of the budding yeast sequence dependent centromere opened up the possibility of similar genetic centromere determination in other species. This was supported by the isolation and sequencing of centromeres from the yeast *Kluyveromyces lactis,* which revealed a similar structure to that found in *S. cerevisiae* (Heus et al., 1990, 1993). This was comprised of a conserved four element structure with two elements (KlCDEI and KlCDEIII) homologous to *S. cerevisiae* centromeric motifs (CDEI and CDEIII, respectively) and an AT-rich element KlCDEII similar to *S. cerevisiae* CDEII. An additional adjacent element, KlCDE0, was also identified. Despite these similarities, *K. lactis* centromere fragments do not confer functional centromere activity in *S. cerevisiae* (Heus et al., 1990). This species specificity is partially mediated by CDEIII/KlCDEIII (Heus et al., 1994). Substituting *S. cerevisiae* CDEIII for KlCDEIII in a *K. lactis* centromere fragment confers partial centromere functionality in *S. cerevisiae* and abolishes function in *K. lactis*. Similar centromere organization was also observed in the pathogenic yeast *Candida glabrata* (Kitada et al., 1996).

Chromosome fusions can lead to the formation of dicentric chromosomes containing two canonical centromere sequences. Consistent with the presence of sequence dependent centromere identity, budding yeast dicentric chromosomes that were generated by integrating centromeric sequences into a chromosomal arm resulted in low viability due to the presence of two functional centromeres, but some cells survive through deletion of the canonical centromere via flanking mobile elements (Surosky and Tye, 1985). Alternatively, the problems associated with dicentric chromosomes can be resolved through chromosome breakage between the two centromeres and subsequent telomere repair on the chromosome fragments (Jäger and Philippsen, 1989). Dicentric chromosome formation induced using a conditional centromere that is supressed by transcription from an adjacent regulatable promoter leads to dicentric chromosome formation and chromosome breakage (Hill and Bloom, 1987, 1989). Dicentric minichromosomes can be propagated through cell division, albeit unstably and the stability of dicentric minichromosomes increases when the two centromeres are located more closely together or following mutations

in CDEII and CDEIII that affect the efficiency of one the centromeres (Koshland et al., 1987; Sullivan and Willard, 1998).

Epigenetic regional mammalian centromeres

Mammalian centromeres form over regional domains and are thus termed regional centromeres, in contrast to the point centromere of budding yeast. Initial analyses indicated that human centromeres potentially also form in a sequence dependant manner. This idea found early support with the identification of alpha satellite arrays composed of ~171 bp tandem repeats found specifically at all human centromeres with different subclasses of alpha satellite structures found at different centromeres (Figure 1.1; Choo et al., 1991; Mitchell et al., 1985). These arrays have a higher order structure composed of repeat units in an array contain the same diverged set of individual repeats and arrays on particular chromosomes are more related than those of other chromosomes (Waye and Willard, 1989). Centromeric repeats also contain CENP-B box motifs, which are bound by the centromeric protein CENP-B, while non centromeric alphoid repeats lack these characteristics (Ando et al., 2002; Masumoto et al., 1989a; Ohzeki et al., 2002; Willard, 1985). However, a direct relationship between sequence and centromere identity is not present since CENP-B boxes are not necessary for centromere maintenance but have been reported to be necessary (though not sufficient) for *de novo* centromere formation (Ikeno et al., 1998; Masumoto et al., 1989a; Okada et al., 2007). Furthermore, CENP-B boxes are absent from human and chimpanzee Y chromosomes (Haaf et al., 1995) and CENP-B may not be strictly necessary for centromere formation, as discussed below. Moreover, the higher order repeats are not always sufficient for centromere formation. The higher order repeats alone were able to confer centromere function to an artificial human chromosome (Schueler et al., 2001), but centromere formation on alpha satellite sequences of human artificial chromosomes is at least partially dependent on sequence length (Grimes et al., 2002, 2004). Characterisation of centromeres formed on satellite repeats has been hampered by the highly repetitive nature of the sequences, though these limitations are being addressed (Hayden and Willard, 2012).

However, human centromeres were subsequently shown to be determined epigenetically, in part due to the observation of neocentromeres that form on normally non-centromeric sequences. Initial cytogenetic study of a clinical sample

revealed the presence of an additional chromosome derived from chromosome 10, which contained a neocentromere that had formed in a normally euchromatic region and lacked alpha satellite repeats (Voullaire et al., 1993). Further analysis confirmed the absence of alpha satellite repeats on the additional chromosome and showed that the centromeric sequence was unique to that chromosome (Sart et al., 1997). Rather, the neocentromere sequence was found to be AT-rich and enriched for repeat elements distinct from centromeric repeats (Lo et al., 2001). Multiple human neocentromeres have been observed to form at non-alpha satellite sequences and at regions distant from heterochromatin (Alonso et al., 2010; Nakano et al., 2003; Sart et al., 1997). Neocentromeres are associated with centromeric proteins CENP-A and CENP-C, which indicate centromere and kinetochore formation, respectively (Warburton et al., 1997), but many do not contain a CENP-B box or bind CENP-B (Alonso et al., 2010; Chueh et al., 2005; Lam et al., 2006). These neocentromeres are nevertheless able to recruit all canonical centromeric proteins tested and form functional kinetochores (Saffery et al., 2000). Numerous human neocentromeres have been identified (Barry, 1999; Burnside et al., 2011; Marshall et al., 2008; Tyler-Smith et al., 1999). Indeed, this process seems to be important on evolutionary timescales as well since humans possess a number of centromeres that have been repositioned relative to other primate species (Rocchi et al., 2012; Ventura et al., 2007) and human centromeres have been observed to relocate back to ancestral locations (Capozzi et al., 2009; Ventura et al., 2004).

Further evidence for epigenetically defined human centromeres came from observations of dicentric chromosomes. Human dicentric chromosomes produced by Robertsonian translocation often only have one remaining active centromere, marked as the primary chromosomal constriction, while the second centromere is inactivated and does not appear as a similar constriction (Earnshaw et al., 1989; Sullivan and Schwartz, 1995). Inactive centromeres show the presence of CENP-B and alpha satellite repeats by immunostaining, but the kinetochore proteins CENP-C and CENP-E are absent from inactivated centromeres, indicating that a kinetochore does not form at these sites despite the presence of centromeric sequences. Robertsonian translocations can also be induced experimentally through telomere disruption, which produces stable dicentric chromosomes with two functional centromeres, as well as chromosomes with a single functional centromere, resulting from either epigenetic inactivation or excision of alpha satellite sequences

(Stimpson et al., 2010). Dicentric chromosomes have also been produced through the reversible integration of an artificial chromosome containing centromeric alpha satellite sequences into a chromosome arm (Nakano et al., 2003). Centromeric proteins CENP-A and CENP-C were initially absent from the integrated minichromosome, consistent with epigenetic inactivation, but upon excision CENP-A and CENP-C were once again detected upon minichromosome reformation, consistent with centromere re-activation.

Epigenetic regional chicken centromeres

Similar to mammalian centromeres, chicken centromeres form large regional centromeres associated with repeat sequences, though centromeres on chromosomes 5, 27, and Z occur on unique sequences (Shang et al., 2010). Centromere identity is determined epigenetically since neocentromeres in chicken cells can be induced by cre-lox recombinase excision of the chromosome Z and 5 centromeres in DT40 cells (Shang et al., 2013). Although neocentromeres form over the length of the chromosome, there is a bias towards sites around the deleted centromere that may be driven by low-level enrichment of the centromeric protein CENP-A in surrounding regions. CENP-A acts as an epigenetic mark for centromere identity and is discussed in more detail below (Black and Bassett, 2008). There are no defining sequences that determine neocentromere identity, although they preferentially form on AT-rich sequences, and have no enrichment of histone modifications associated with heterochromatin or active transcription. These neocentromeres form in transcriptionally inactive regions, but the Z chromosome only occurs in one copy so disruption of Z chromosome gene expression may be lethal. Chromosome 5 neocentromere can form over transcribed regions, but it is not known whether they remain transcriptionally active following neocentromere formation.

Epigenetic regional Candida albicans centromeres

Candida albicans canonical centromeres form over regional domains with each chromosome possessing a unique centromeric sequence (Figure 1.1; Sanyal et al., 2004). All centromeres except one are flanked by inverted or direct repeat sequences. These repeat sequences are not shared between centromeres on different chromosomes and overall the centromeres are not enriched for repetitive sequences (Mishra et al., 2007). The centromeric sequences themselves are not

sufficient for centromere formation since the introduction of naked DNA containing an entire centromere does not form a functional centromere (Baum et al., 2006). Centromere identity is epigenetic since deletion of a centromere in *C. albicans* allows neocentromeres to form in multiple locations not associated with centromere sequences. Neocentromere were found to form frequently close to the deleted canonical centromere but also in intergenic and subtelomeric regions (Ketel et al., 2009). Despite the range of locations where neocentromeres form, the regions where they arise share some similar attributes, although these conclusions are limited by the relatively small number of neocentromeres analysed. Noticeably, *C. albicans* neocentromeres tend to form near repetitive sequences, similarly to canonical centromeres. At neocentromeres formed in intergenic regions the repeat sequences are composed of flanking genes with shared sequence homology. Furthermore, neocentromeres can be induced to change their localization due to environmental stress (Ketel et al., 2009).

Epigenetic regional fission yeast centromeres

Fission yeast, *Schizosaccharomyces pombe*, centromeres are ~35-115 kb in size and are comprised of a non-repetitive central core flanked by centromere specific inner- and outer-repeat elements (Figure 1.1; Clarke and Baum, 1990; Clarke et al., 1986; Fishel et al., 1988; Hahnenberger et al., 1991; Murakami et al., 1991; Nakaseko et al., 1987). The central core is immediately flanked by the inverted inner-repeats that are unique to each centromere (Clarke and Baum, 1990; Hahnenberger et al., 1991; Murakami et al., 1991). The outer-repeat elements are ~4.4 kb and ~4.8 kb in length and termed dg and dh (or L and M) repeats, respectively (Clarke et al., 1986; Nakaseko et al., 1987). These repeat elements are present at all centromeres in a variable number and form a chromatin environment that exhibits repressed recombination and transcription (Clarke and Baum, 1990; Clarke et al., 1986; Fishel et al., 1988; Hahnenberger et al., 1991; Murakami et al., 1991; Nakaseko et al., 1987). Two distinct domains of chromatin are formed at fission yeast centromeres (Partridge et al., 2000). The central core and the proximal part of the inner-repeats (collectively termed the central domain) are enriched for centromeric proteins associated with the kinetochore, while the distal part of the inner-repeats and the outer-repeats form heterochromatin and associate with the heterochromatic proteins Swi6 (homologous to HP1) and Chp1. These two domains are separated by tRNA boundary elements. Genes inserted into the central domain

and outer repeats are transcriptionally silenced through different mechanisms (Allshire et al., 1994, 1995; Partridge et al., 2000). The central core is necessary for centromere activity on a minichromosome, but there is no specific sequence within the central core that is itself required and the central core alone is unable to direct kinetochore assembly (Baum et al., 1994; Catania et al., 2015). Neighbouring heterochromatin, directed by outer repeats or by artificial tethering, is required to induce kinetochore assembly on the central domain (Folco et al., 2008; Kagansky et al., 2009).

Centromere identity in *S. pombe* is determined epigenetically. Neocentromeres have been generated in *S. pombe* through cre-lox recombination mediated centromere deletion (Ishii et al., 2008). Neocentromeres the centromere proteins CENP-A^{Cnp1} and CENP-C^{Cnp3} occupy a regional domain relatively close to a heterochromatic domain similar to canonical centromeres. These neocentromeres form over a domain of low transcription of tens of kilobases proximal to the subtelomeric heterochromatin in a manner that is only partially dependant on heterochromatin. Furthermore, neocentromeres were also found to form internal to the arrays of subtelomeric rDNA repeats on chromosome III (Ogiyama et al., 2013). However, such neocentromeres show lower levels of CENP-A^{Cnp1} and its chaperone Scm3 and are unstably propagated. Stable chromosome segregation was recovered upon fusion with another chromosome carrying a normal centromere, or a rearrangement that brought the neocentromere closer to the telomere. Dicentric chromosomes were also created by forced recombination using cre/lox or selectable markers (Sato et al., 2012). Dicentric chromosomes were resolved by chromosome breakage, centromere deletion, or irreversible epigenetic inactivation. Centromere inactivation was associated with histone H3 hypoacetylation and heterochromatin formation across the central domain of one centromere. However, this inactivation did not require heterochromatin factors and in the presence of heterochromatin this epigenetic centromere inactivation was irreversible, even upon deletion of the active centromere.

Holocentromeres

Despite large organisational differences, point and regional centromeres both form at a single locus on chromosomes. Such chromosomes are termed monocentric. In contrast, a third type of centromeres, termed holocentromeres, form along the entire

length of chromosomes and have arisen independently multiple times in plants and animals (Melters et al., 2012). Holocentric chromosomes can be identified in a number of ways: mitotic holocentric chromosomes lack a primary constriction; chromatids segregate parallel to spindle poles; following chromosome fragmentation individual fragments retain segregation activity; and centromere proteins are enriched across the length of the chromosomes (Melters et al., 2012). The most well studied holocentric species is the nematode *Caenorhabditis elegans.* In *C. elegans* the centromeric protein CeCENP-A is enriched at regional domains located across the length of chromosomes and in embryos is correlated with regions that are transcriptionally inactive in the maternal germline (Gassmann et al., 2012). A later study found evidence that there are additional peaks of CeCENP-A primarily in intergenic regions that resemble point centromeres, which were proposed to be the primary sites of kinetochore formation (Steiner and Henikoff, 2014). There may be changes in the genome that associate with transition to holocentricity. In insects four independent transitions to holocentricity were correlated with the loss of the centromeric protein CENP-A in each case and it was proposed that some change in centromere organisation or kinetochore structure made these species susceptible to transitioning to holocentricity with an associated loss of CENP-A, but loss of CENP-A is not observed in other holocentric species (Drinnenberg et al., 2014). A similar transition from monocentricity to holocentricity may even be in the process of occurring in the pea *Pisum sativum* and related species (Neumann et al., 2012, 2015). These species contain metapolycentric chromosomes that possess elongated primary constrictions that span up to hundreds of megabases and in some cases are associated with multiple domains of CENP-A (Neumann et al., 2012, 2015).

Evolutionary new centromeres

Neocentromeres that become fixed in the population can then segregate as evolutionary new centromeres (ENCs). As mentioned above, chicken chromosomes 5, 27, and Z have non-repetitive centromeres and were proposed to be ENCs, (Shang et al., 2010). An ENC has also been identified on chromosome 11 in the domesticated horse *Equus caballus* (Wade et al., 2009). This horse ENC does not contain alpha satellite sequences or other repetitive sequences normally associated with mammalian centromeres. Sequencing of cell lines from multiple horses revealed distinct chromosomal domains on chromosome 11 coated with the CENP-

A in different individuals. Moreover, some individuals contained two non-overlapping domains of CENP-A (Purgato et al., 2014). Single nucleotide polymorphisms showed that in those individuals with two CENP-A domains the two homologous chromosome 11s carried different singular CENP-A domains. ENCs have also been observed in zebra, horse, donkey, humans and macaques (Carbone et al., 2006; Ventura et al., 2007). Indeed, there may be a general evolutionary trajectory for ENCs where they initially form on non-repetitive sequences, as observed for human neocentromeres and chicken and horse ENCs, and later accumulate repeat elements, as observed for macaque ENCs (Figure 1.2; Ventura et al., 2007).

It is clear that centromere identity and organisation can vary widely, from a sequence dependent point centromere to epigenetically regulated regional centromeres and holocentromeres. Nevertheless, some of the centromere and kinetochore proteins are conserved in diverse species.

1.2 Centromere and kinetochore proteins

Studies of vertebrate kinetochore structure have contributed greatly to the understanding of kinetochore structure. Electron microscopy of HeLa cells showed that multiple spindle microtubules make attachments with a single kinetochore (Robbins and Gonatas, 1964). The kinetochore itself forms a trilaminar plate, composed of electron dense inner- and outer-plates separated by an electron sparse middle plate (Comings and Okada, 1971; Jokelainen, 1967; Roos, 1973). The interaction between DNA and the kinetochore is restricted to the inner-plate, while no DNA was observed in association with the outer plate in muntjac by immunoelectron microscopy (Cooke, 1993). Kinetochore structure and centromere organisation vary somewhat between species, though some kinetochore and centromere proteins are conserved (Table 1.1; adapted from Westermann and Schleiffer, 2013).

CENP-A

CENP-A is a centromere specific histone H3 variant (Figure 1.3). Human CENP-A is a 17-kD protein that is centromere specific and forms a MNase resistant particle that is associated with canonical histone proteins (Palmer et al., 1987). Isolation of CENP-A from bull sperm revealed that CENP-A has homology to the canonical histone H3 (Palmer et al., 1991). Human CENP-A similarly has homology to histone

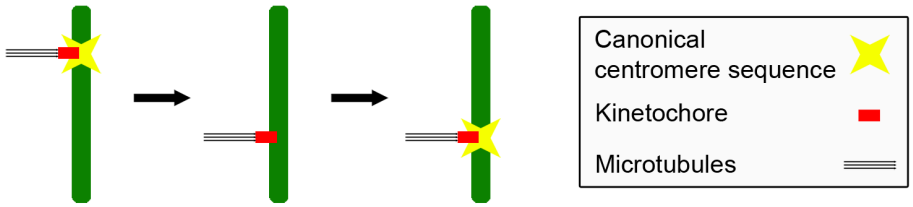
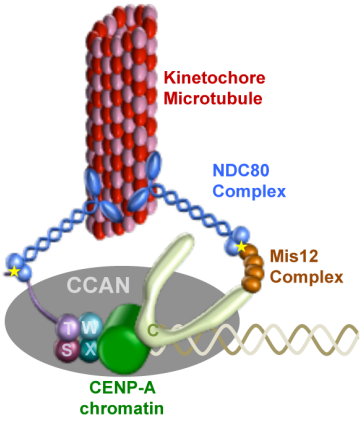


Figure 1.2. Neocentromere formation and progression of evolutionary new centromeres. Evolutionary new centromeres (ENCs) are thought to follow a general evolutionary progression in species with epigenetic centromere identity. Initially the kinetochore forms at sites associated with canonical centromere sequences. Deletion of canonical centromere sequences can lead to neocentromere formation in which kinetochore formation and microtubule attachments occur at a region not associated with centromere sequences. If this neocentromere becomes fixed in the population then it can segregate as an ENC. Centromeric sequences subsequently accumulate at the ENC over evolutionary timescales so that the centromere is once again associated with canonical centromere sequences, suggesting that some aspect of centromeric sequences is important but not essential to centromere function.

Constitutive Centromere Associated Network



CCAN	Mis6C	Ctf19C
Human	*S.pombe*	*S. cerevisiae*
CENP-C	Cnp3	Mif2
CENP-H	Fta3	Mcm16
CENP-I	Mis6	Ctf3
CENP-K	Sim4	Mcm22
CENP-L	Fta1	Iml3
CENP-M	-	-
CENP-N	Mis15	Chl4
CENP-O	Mal2	Mcm21
CENP-P	Fta2	Ctf19
CENP-Q	Fta7	Okp1
CENP-R	-	-
CENP-U	Mis17	Ame1
CENP-S	Mhf1	Mhf1
CENP-X	Mhf2	Mhf2
CENP-T	Cnp20	Cnn1
CENP-W	New1	Wip1
-	Fta4	Nkp1
-	Cnl2	Nkp2
-	Fta5	-
-	Fta6	-

Table 1.1. Homologs of constitutive centromere associated network components (adapted from Westermann and Schleiffer, 2013). Shown here are the human components of the constitutive centromere associated network (CCAN) and their homologs in *S. pombe* and *S. cerevisiae*.

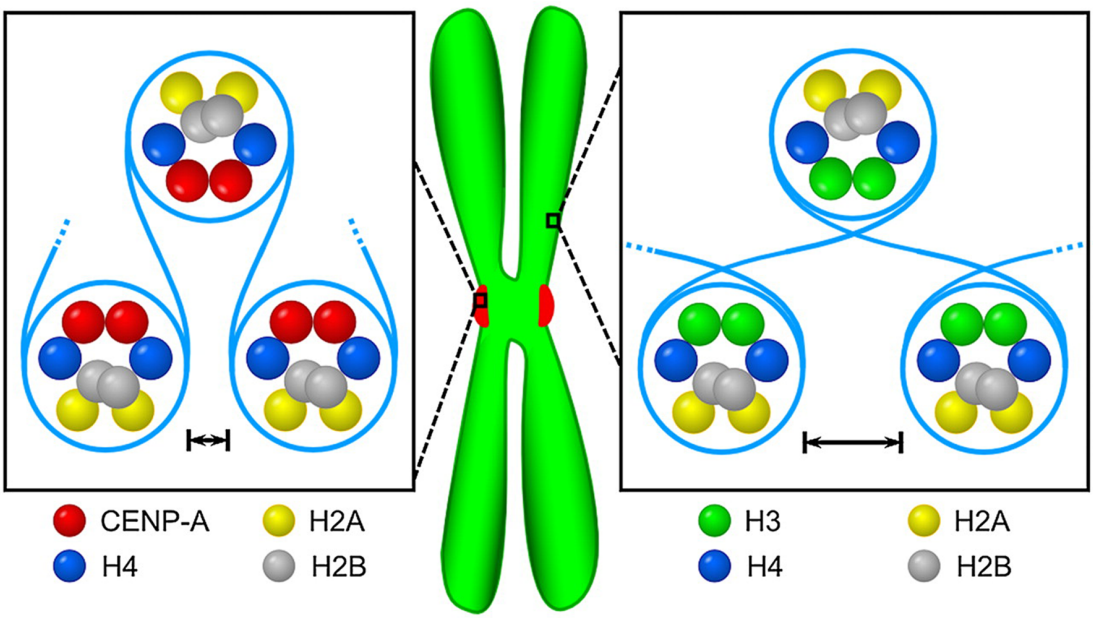


Figure 1.3. CENP-A in a centromere specific histone H3 variant (Panchenko et al. 2011). CENP-A replaces histone H3 and forms an octameric complex with two copies each of histones H2A, H2B and H4. CENP-A nucleosomes are found specifically at the centromere.

H3 (Sullivan et al., 1994) and purified CENP-A can replace H3 in an octameric complex with H2A, H2B, and H4 *in vitro* (Yoda et al., 2000). Thus CENP-A can replace histone H3 to form distinct types of nucleosomes. The homology with H3 is restricted to the C-terminus, which contains the histone fold domain, while the N-terminal domain is variable between species. The C-terminal histone fold domain was shown to confer centromere localization in chimeric CENP-A/H3 proteins, while the unique N-terminal domain did not.

CENP-A, along with CENP-B and CENP-C, localises to centromeres constitutively throughout the cell cycle (Earnshaw and Rothfield, 1985; Masumoto et al., 1989b). *In vivo* CENP-A was shown to be mainly associated with alpha satellite DNA (Vafa and Sullivan, 1997). However, it is only associated with a subset of centromeric alpha satellite DNA types and at neocentromeres CENP-A chromatin assembles on non-alphoid DNA (Warburton et al., 1997). All other proteins that associate with centromeres throughout the cell cycle, termed the constitutive centromere associated network (CCAN), function downstream of CENP-A and require CENP-A for their proper localization (Carroll et al., 2010; Hori et al., 2008). CENP-A is discussed further below.

CENP-B

Human CENP-B binds to a 17 bp motif, termed the CENP-B box, present between CENP-A nucleosomes at some alpha satellite repeats (Henikoff et al., 2015; Masumoto et al., 1989a; Muro et al., 1992). CENP-B may position flanking CENP-A nucleosomes. *In vitro* studies showed that CENP-B binds to DNA as a dimer via N-terminal alpha helices (Yoda et al., 1992). The crystal structure revealed that these helices form two helix-turn-helix domains that induce kink-straight-kink bends in CENP-B box DNA (Tanaka et al., 2001). CENP-B dimers can bind to two CENP-B boxes on a DNA strand to position a nucleosome on the intervening DNA loop (Yoda et al., 1998). ChIP analyses also demonstrate that CENP-A containing nucleosomes are positioned *in vivo* around CENP-B boxes (Ando et al., 2002). Unlike other centromeric proteins such as CENP-A and CENP-C, CENP-B can be detected at inactivated centromeres on dicentric chromosomes by immunostaining

of mitotic spreads, consistent with sequence dependent binding to CENP-B boxes (Earnshaw et al., 1989).

The CENP-B box is also present in mouse (*Mus musculus*) centromeric minor satellite repeats and similarly co-localises with CENP-B by immunostaining (Mitchell et al., 1993). Full CENP-B box motifs are not present in *Mus caroli* but a degenerate centromeric motif containing the nine essential base pairs of a CENP-B box are detectable and can bind both mouse and human CENP-B (Kipling et al., 1995). African green monkey alpha satellite repeats lack CENP-B boxes despite the presence of a CENP-B homolog (Goldberg et al., 1996). Fission yeast contains three potential CENP-B homologs: the replication origin binding protein Abp1 (Locovei et al., 2006), the replication origin binding protein Cbh1 that also binds to centromeric repeat sequences in vitro (Lee et al., 1997), and the centromere central core and inner repeat associated protein Cbh2 (Irelan et al., 2001). Abp1 and Cbh1 are required for transcriptional silencing and heterochromatin structure at the centromeric outer-repeats (Nakagawa et al., 2002). However, human CENP-B and the fission yeast proposed CENP-B homologs belong to different clades of pogo-like transposases so it is likely that the three fission yeast proteins are not homologs of CENP-B, but rather these arose from the convergent evolution of pogo-like transposases (Casola et al., 2008).

CENP-C and the Mis12 complex

Interactions between centromeric CENP-A chromatin and the outer kinetochore are mediated via two pathways, one of which is via CENP-C and the Mis12 complex. CENP-C localises to the centromere via a 60 bp domain and can bind DNA non-specifically via a distinct domain (Yang et al., 1996). Like human CENP-C, the budding yeast CENP-C^{Mif2} is a centromere specific protein and contains a CENP-C box and AT-hook that are required for centromere localization (Cohen et al., 2008; Meluh and Koshland, 1995, 1997). Immuno-electron microscopy localised human CENP-C to a thin band associated only with the kinetochore inner-plate and not reflecting the broader distribution of CENP-B, which additionally localises to centromeric heterochromatin (Saitoh et al., 1992).

Most evidence supports an interaction between CENP-A and CENP-A nucleosomes. Initial co-immunoprecipitation results indicated that CENP-C directly associated with

CENP-A (Ando et al., 2002). *In vitro* CENP-C was found to bind directly to DNA and was also shown to bind the extreme C-terminal region of CENP-A (Carroll et al., 2010). CENP-C interacts directly with CENP-A and CENP-C recruitment to endogenous centromeres is dependent on CENP-A (Carroll et al., 2010; Fukagawa et al., 2001; Hori et al., 2008; Kwon et al., 2007; Milks et al., 2009). CENP-C alone is unable to recruit other centromeric proteins following over-expression so cannot itself mediate functional centromere formation (Fukagawa et al., 1999). However, tethering human CENP-C to an ectopic locus is sufficient to assemble functional kinetochores (Gascoigne et al., 2011; Przewloka et al., 2011; Rago et al., 2015). *In vivo* the C-terminal Mif2 homology domain (named for the *S. cerevisiae* homolog) is necessary for proper centromere localization and also confers proper localization of human CENP-C in *Xenopus* cells (Lanini and McKeon, 1995). However, in chicken DT40 cells the co-immunoprecipitation of CENP-C with CENP-A disappeared after further MNase digestion, suggesting that CENP-A and CENP-C do not directly interact, and CENP-C co-immunoprecipitated with H3 under these conditions (Hori et al., 2008). This difference might be due to disrupting the interaction between CENP-C and CENP-A under these conditions.

CENP-C interacts with the Mis12 complex and subsequently the outer kinetochore Ndc80 complex. The Mis12 complex contains 4 subunits: known as Mtw1/Mis12, Dsn1/Mis13, Nsl1/Mis14, and Nnf1 depending on the organism being studied (Euskirchen, 2002; Kline et al., 2006; Pinsky et al., 2003). Human and *Drosophila* Mis12 has also been shown to associate with the N-terminus of CENP-C. The *in vitro* reconstituted human Mis12 complex has been shown to associate with the Ndc80 complex via the C-terminus of Nsl1 (Petrovic et al., 2010) and *Drosophila* Ndc80 requires Mis12 and Nsl1 for localisation to centromeres (Venkei et al., 2011). Thus the Mis12 complex appears to form a conserved bridge between the inner- and outer-kinetochore complexes connecting the kinetochore with microtubules via its association with Ndc80 and with chromatin via CENP-C (Figure 1.4; Screpanti et al., 2011). Consistent with this, depletion of CENP-C in HeLa and DT40 cells leads to kinetochore disruption and mitotic death (Fukagawa and Brown, 1997; Tomkiel et al., 1994). Similarly, *S. cerevisiae* CENP-C^{Mif2} interacts with CENP-A^{Cse4} and *CENP-*C^{mif2} mutations lead to defects in attachments between chromosomes and spindle microtubules (Pinsky et al., 2003). Conditional CENP-C deletion in DT40 cells reveals that CENP-A localisation at centromeres is unaffected by loss of CENP-C,

but the Mis12 complex shows highly reduced centromere association (Kwon et al., 2007). However, interpreting the role of Mis12 was complicated by the finding that artificial tethering of human CENP-T also recruits Mis12 and allows formation of functional kinetochores (Rago et al., 2015). Moreover, at endogenous centromeres the Mis12 complex was found to be recruited by CENP-T, even in the absence of CENP-C, and Mis12 recruitment by CENP-T was shown to require the Ndc80 complex (Rago et al., 2015). Thus, it was proposed that the CENP-C and CENP-T pathways mediate interactions between chromatin and the outer kinetochore are not redundant, but rather specifically regulated aspects of a more complex architecture.

CENP-T/W/S/X Complex

The second pathway that mediates the interaction between centromeric CENP-A chromatin and the outer kinetochore is via the CENP-T/W/S/X complex. CENP-T was also identified as a CCAN component that localises to centromeres throughout the cell cycle (Figure 1.4; Foltz et al., 2006; Obuse et al., 2004). Immuno-electron microscopy showed that CENP-T localised to the inner kinetochore of DT40 chromosomes (Suzuki et al., 2011). CENP-T was identified by immunoprecipitation from chicken DT40 cells followed by mass spectrometry as forming a complex with CENP-W (Hori et al., 2008). Both CENP-T and CENP-W are essential proteins that contain histone fold domain and their localisation to centromeres displays interdependency and is dependent on their histone fold domains and CENP-A (Hori et al., 2008). However, in *Xenopus* egg extracts it was found that CENP-T and CENP-W display distinct temporal localisation patterns and depletion of CENP-W only partially co-depleted CENP-T but also co-depleted CENP-C, suggesting that CENP-T and CENP-W form distinct complexes in addition to their direct interaction (Krizaic et al., 2015).

CENP-S and CENP-X were identified in immunoprecipitates of Fanconi anemia nuclear core complex (FANC) from HeLa cells by mass spectrometry and originally named Mhf1 and Mhf2, respectively (Singh et al., 2010; Yan et al., 2010). CENP-S/Mhf1 and CENP-X/Mhf2 interact more closely with each other than with FANC and form a heterodimer capable of binding double stranded DNA (Yan et al., 2010). Structural analyses have revealed that CENP-S/CENP-X form a hetero-tetramer that binds to a single subunit of FANCM (Tao et al., 2012). *In vitro* the CENP-S/CENP-X complex cooperatively binds DNA and in HeLa cells they associate with

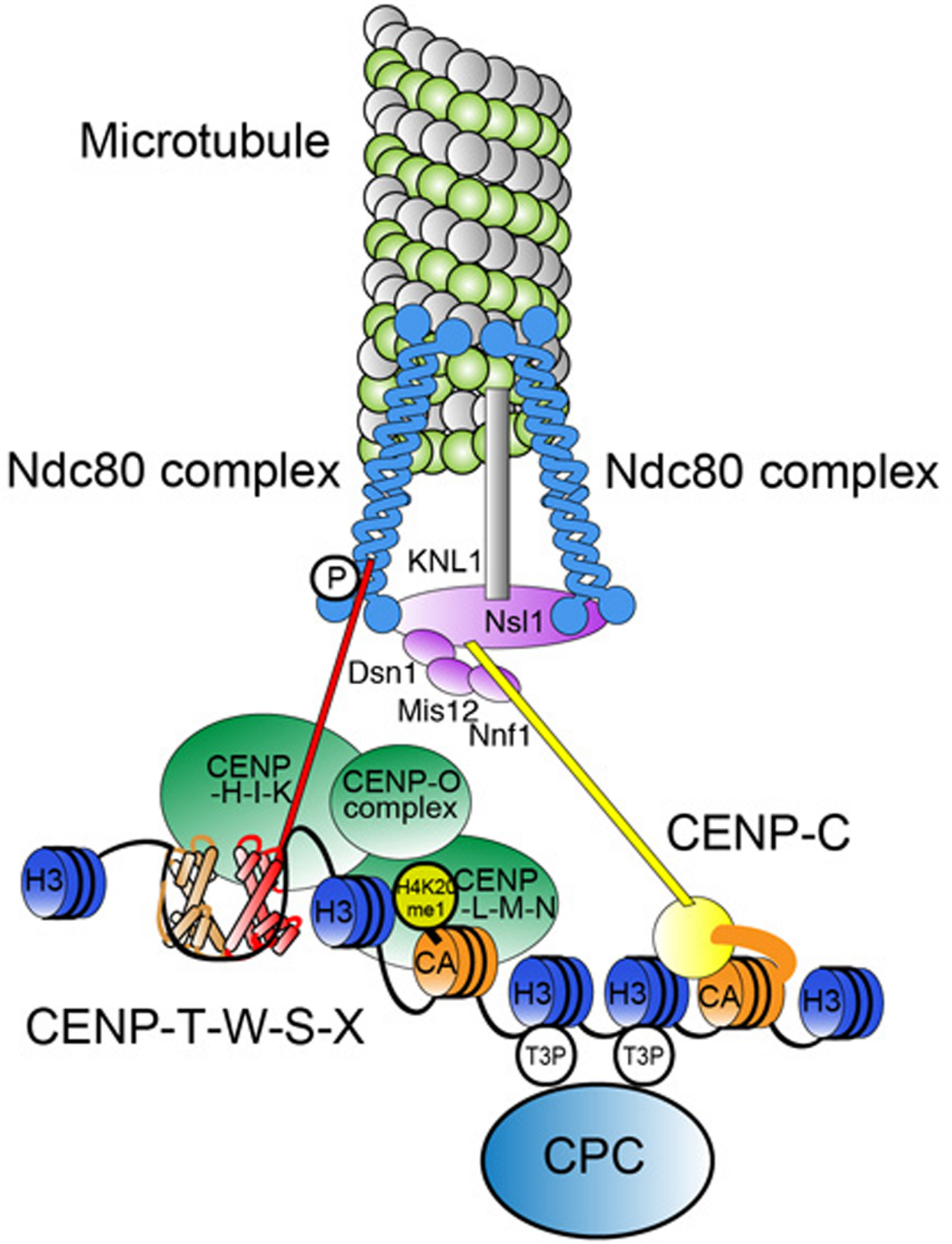


Figure 1.4. Kinetochore structure and centromeric attachments (Fukagawa and Earnshaw 2014). There are two redundant pathways connecting centromeric chromatin with the Ndc80 complex, which interacts with microtubules. One pathway is via the CENP-T/W/S/X complex, which binds DNA or interacts with centromeric nucleosomes and directly interacts with the Ndc80 complex. The other pathways is

DNA crosslinks induced during S phase (Yan et al., 2010). The fission yeast homologue of the FANCM homolog is Fml1 (Yan et al., 2010). Although CENP-S and CENP-X (Mhf1 and Mhf2) also associate with Fml1 in fission yeast, deletion of the *fml1* gene does not affect chromosome segregation, while cells lacking CENP-S^{Mhf1} or CENP-X^{Mhf2} exhibit defective chromosome segregation (Bhattacharjee et al., 2013). This indicates that CENP-S^{Mhf1} and CENP-X^{Mhf2} have a function at centromeres outside of its interaction with Fml1. Consistent with this Fml1 is required for Mhf1/Mhf2 localization to sites of DNA repair and recombination, but not for their localization to centromeres.

The vertebrate CENP-T/W/S/X complex may interact with CENP-A or H3 nucleosomes (Foltz et al., 2006a; Ribeiro et al., 2010; Schleiffer et al., 2012) but it was also proposed to form a nucleosome-like DNA binding complex (Nishino et al., 2012). Super-resolution microscopy of unwound chromatin fibers from chicken DT40 cells revealed that CENP-T localises to the H3 domains interspersed between CENP-A nucleosomes (Ribeiro et al., 2010). Furthermore, FRET experiments using human CENP-T found that CENP-T resides in close proximity to H3.1 but not H3.2, H3.3, or CENP-A (Abendroth et al., 2015). Earlier reported interactions between CENP-A and CENP-T may have resulted from partial MNase digestion of chromatin (Foltz et al., 2006a). Nevertheless, CENP-T complex is recruited to centromeres via CENP-A and associated proteins (Fukagawa et al., 2001; Okada et al., 2006; Sugata et al., 1999).Interestingly *Drosophila* lacks proteins homologous to CENP-T, -W, -S, and -X, along with counterparts of many other kinetochore components that are conserved between other metazoans (Westermann and Schleiffer, 2013). It therefore relies completely on the CENP-C-Mis12-Ndc80 pathway for kinetochore-microtubule attachment.

CENP-T, CENP-S, and CENP-X were shown to localise to centromeres during S and G2, but CENP-T is not propagated across cell cycles and is assembled every cell cycle (Dornblut et al., 2014; Prendergast et al., 2011). This suggests that if the CENP-T/W/S/X complex binds to DNA as a nucleosome-like complex then it must be assembled every cell cycle unlike the epigenetic mark CENP-A. CENP-T directly interacts with the outer kinetochore Ndc80 complex and plays a role in anchoring the outer kinetochore (Figure 1.4; Schleiffer et al., 2012). Tethering of CENP-T can also induce ectopic kinetochore formation in humans and budding yeast, but does

not recruit CENP-A during kinetochore formation (Gascoigne et al., 2011; Rago et al., 2015; Schleiffer et al., 2012).

The Ndc80 complex

Both the CENP-T/W/S/X and CENP-C/Mis12 complex pathways ultimately interact with the outer kinetochore Ndc80 complex. The Ndc80 complex contains four conserved subunits: Ndc80/Hec1, Nuf2, Spc24, and Spc25 (Janke et al., 2001; McCleland et al., 2004; Wigge and Kilmartin, 2001). EM analysis of the budding yeast Ndc80 complex found that two subcomplexes, Ndc80/Nuf2 and Scp24/Spc25, form a tetrameric rod-like structure with the two subcomplexes connected by coiled-coil domains and globular heads at both ends of the complex (Wei et al., 2005). The Spc24/Spc25 dimer is responsible for the localisation of the Ndc80 complex, while the Ndc80/Nuf2 dimer interacts with microtubule ends (Ciferri et al., 2008). The Spc24/Spc25 dimer associates directly with both $Mis12^{Mtw1}$ and $CENP\text{-}T^{Cnn1}$ (Petrovic et al., 2010; Schleiffer et al., 2012). $Mis12^{Mtw1}$ and $CENP\text{-}T^{Cnn1}$ share a similar motif that is required for competitive binding to the same site in the Spc24/Spc25 dimer (Malvezzi et al., 2013). EM tomography of vertebrate kinetochores showed that fibers, which were hypothesised to be the Ndc80 complex, extend from the outer kinetochore plate and interact with microtubules (Dong et al., 2007).

The Ndc80 complex binds to microtubules *in vitro* (Cheeseman et al., 2006). Furthermore, *in vitro* studies showed that *S. cerevisiae* Ndc80 complex coupled to beads allow their association with growing and shortening microtubule ends. This association was moderately maintained even when physical tension was applied under low stringency conditions (Powers et al., 2009). The fission yeast Ndc80 complex is part of a greater NDC80-MIND-Spc7 (NMS) complex (Liu et al., 2005). NMS remains associated with centromeres throughout the cell cycle and also interacts with the spindle pole body (Hayashi et al., 2006; Kerres et al., 2007; Liu et al., 2005; Pidoux et al., 2003). There is evidence of independent Ndc80 and MIND complexes in budding yeast (De Wulf et al., 2003), though the interaction between the two seems to be tighter in fission yeast (Kerres et al., 2007).

Dam1/DASH complex

Interactions between the Ndc80 complex and spindle microtubules are stabilised by other complexes, including the Dam1/DASH complex. The Dam1/DASH complex is composed of subunits: Dam1, Duo1, Ask1, Spc34, Spc14, Dad1, Dad2/Hsk1, Dad4/Hsk2, and Hsk3 (Cheeseman et al., 2001a, 2001b; Janke et al., 2002; Li et al., 2002). Duo1 and Dam1 are essential in budding yeast and mutations exhibit spindle defects and lead to activation of the spindle checkpoint (Cheeseman et al., 2001a). The budding yeast DASH complex can bind to microtubules *in vitro* (Asbury et al., 2006; Cheeseman et al., 2001b) and forms a ring around microtubules that stabilises the microtubule plus end (Figure 1.5; Westermann et al., 2005). Consistent with its proposed ring structure, microscopic analyses suggested that the number of DASH complexes per centromere *in vivo* is sufficient for ring formation in both fission and budding yeast (Coffman et al., 2011). Tethering of Dam1 increases the rate of successful minichromosome segregation in a manner that does not depend in CENP-A^{Cse4} or Ndc80 (Kiermaier et al., 2009; Lacefield et al., 2009). The DASH complex associates with kinetochore microtubule ends *in vivo* and remains associated with lengthening and shortening microtubule ends *in vitro* (Lampert et al., 2010). The DASH complex alone can stably associate with depolymerising microtubule plus ends *in vitro*. In contrast, *in vitro* the Ndc80 complex requires the DASH complex to stabilise interactions with depolymerising microtubules under physiological conditions (Lampert et al., 2010; Tien et al., 2010). The DASH complex also increases the tension bearing capacity of bead-coupled Ndc80 complexes (Tien et al., 2010). The fission yeast DASH complex only localises to the centromere during mitosis and is not essential but its loss leads to errors in chromosome segregation (Liu et al., 2005; Sanchez-Perez et al., 2005). Dad1 is part of both the DASH and Sim4 complexes and is required for the localization of all other DASH components to centromeres (Liu et al., 2005). Dad1 is the only DASH component that localises to fission yeast centromeres throughout the cell cycle (Sanchez-Perez et al., 2005).

The DASH complex is not conserved outside of fungi, although functionally equivalent complexes have been proposed. Vertebrate kinetochore have no known

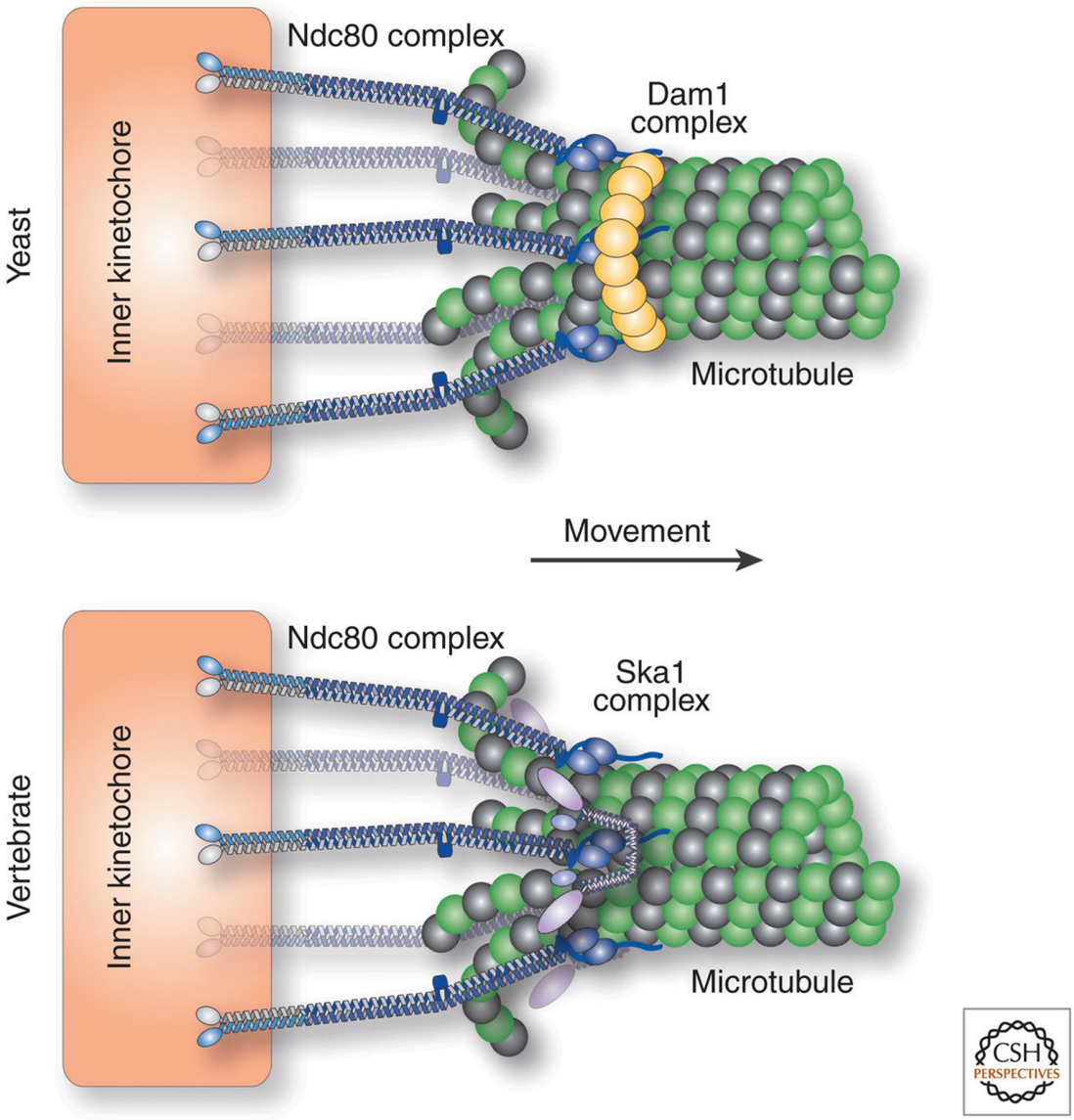


Figure 1.5. Association between between the outer-kinetochore and microtubules (Cheeseman, 2014). The outer-kinetochore component Ndc80 interacts with microtubules, but does not stay associated with depolymerising microtubule ends. In yeast this interaction is mediated by the Dam1/DASH complex, which forms a ring around mictrotubules to increase the load bearing capacity of Ndc80 and allow it to associate with depolymerising microtubule ends. The Dam1/DASH complex is not conserved in vertebrates. Instead, the Ska1 complex may be the functional equivalent to the Dam1/DASH complex. The Ska1 complex can form a ring structure that binds depolymerising microtubule ends, but its role is less well understood.

DASH homolog, but the SKA complex has also been shown to form a ring structure and to bind depolymerizing microtubule plus ends in vitro (Figure 1.5; Gaitanos et al., 2009; Jeyaprakash et al., 2012). However, the SKA complex has not yet been shown to mediate the association of the Ndc80 complex with microtubules.

The above summary demonstrates that some centromere and kinetochore proteins are widely conserved. One of the most widely conserved centromere proteins is CENP-A, which has been shown in many systems to be vital for centromere identity, kinetochore assembly, and thus chromosome segregation.

1.3 CENP-A nucleosomes: the epigenetic mark for centromere identity

In several species analysed it became clear that centromere identity was not dependent on particular DNA sequences and therefore an epigenetic mechanism for determining centromere identity was proposed (Karpen and Allshire, 1997). The centromere specific histone-fold containing protein CENP-A has attracted considerable attention and has become accepted as the epigenetic mark that governs centromere identity (Black and Bassett, 2008). CENP-A localizes to centromeres in all characterized model species and replaces canonical histone H3 in centromere specific nucleosomes (Foltz et al., 2006a; Van Hooser et al., 2001; Yoda et al., 2000). CENP-A nucleosomes are assembled at neocentromeres activated in chromosomal regions with no homology to canonical centromere sequences and CENP-A remains at the active centromere of dicentric chromosomes while it is lost from epigenetically inactivated centromeres (Burrack and Berman, 2012; Sato and Saitoh, 2013; Scott and Sullivan, 2014). Furthermore, the tethering of CENP-A to an ectopic site has been shown to induce functional kinetochore formation at non-centromeric locations on *Drosophila* (Heun et al., 2006; Mendiburo et al., 2011) and human chromosomes (Tachiwana et al., 2015). In addition, tethering of HJURP, the CENP-A chaperone, to an ectopic non-centromeric chromosomal location leads to the incorporation of CENP-A and the formation of a functional kinetochore (Barnhart et al., 2011).

CENP-A evolution

CENP-A is highly functionally conserved with described orthologs in diverse species from yeasts to mammals (Figure 1.6; Blower and Karpen, 2001; Palmer et al., 1987, 1990, 1991; Sanyal and Carbon, 2002; Stoler et al., 1995; Takahashi et al., 2000;

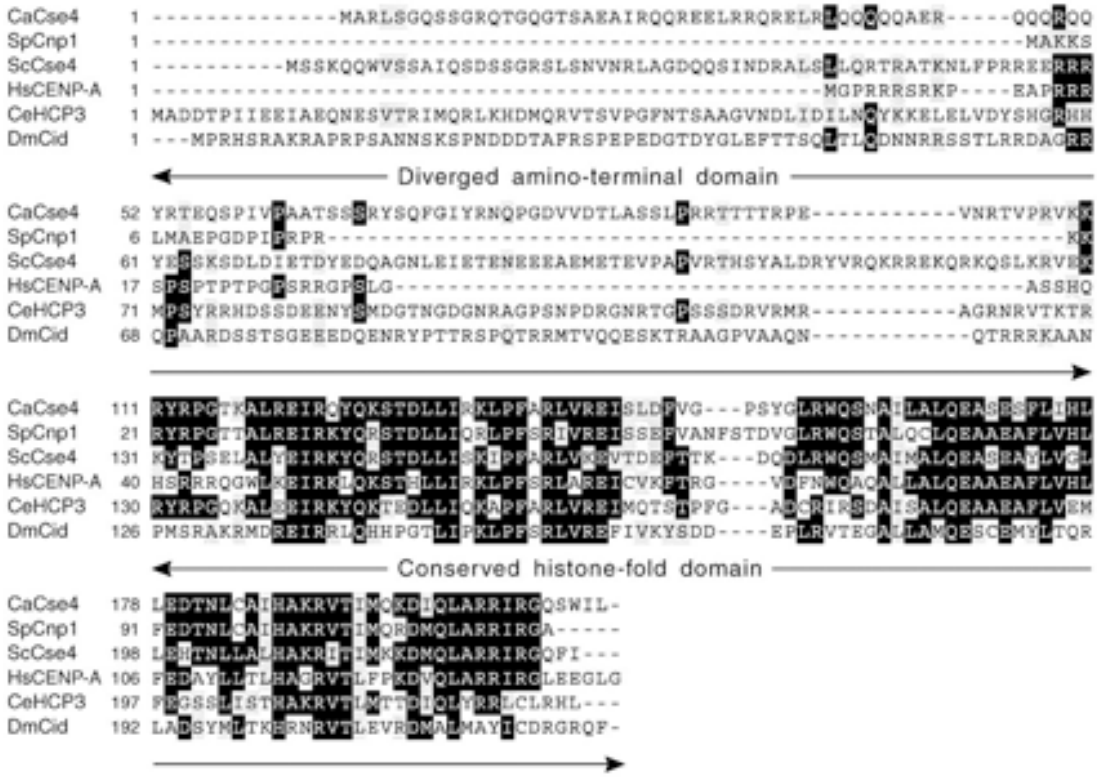


Figure 1.6. Sequence alignment of CENP-A orthologs (Sanyal and Carbon, 2002). Sequence alignments of CENP-A orthologs from *Candida albicans, Schizosaccharomyces pombe, Saccharomyces cerevisiae, Homo sapies, Caenorhabditis elegans,* and *Drosophila melanogaster* are shown. Similar amino acids are indicated by gray boxes and identical amino acids are indicated by black boxes.

Talbert et al., 2002; Wieland et al., 2004). However, CENP-A has undergone rapid evolution in many lineages.

The *Drosophila* CENP-A homolog, CENP-A^{CID}, shows evidence of rapid evolution and positive selection at the N-terminal tail and histone fold domain Loop 1, which are predicted to interact with DNA (Figure 1.7; Malik et al., 2002). CENP-A^{CID} shows length variability between species, which is primarily driven by differential oligorepeat expansions in the N-terminal tail. In contrast, the region of the histone fold domain that is predicted to dimerise with a second subunit is highly conserved. It has been proposed that the adaptive evolution of CENP-A^{CID} is related to the rapid divergence of centromeric satellite repeats between *Drosophila* species (Malik and Henikoff, 2001). Indeed, when *D. bipectinata* CENP-A^{CID} was expressed in *D. melanogaster* it was found to be unable to localize to *D. melanogaster* centromeres due to divergence in the histone fold domain loop 1 (Vermaak et al., 2002). The centre of loop 1 was found to be unnecessary for targeting to centromeres. Further analyses revealed that regions on either end of loop 1 of CENP-A^{CID}, as well as the increased length of CENP-A^{CID} loop 1 relative to histone H3 loop 1, were required for its targeting to centromeres. Similarly, *Arabidopsis thaliana* CENP-A^{cenH3} has a conserved histone fold domain with high similarity with H3, but the N-terminal tail shares no similarity to histone H3 and has undergone rapid positive selection (Talbert et al., 2002). In primates CENP-B is highly conserved with no evidence of positive selection, while CENP-A and CENP-C are rapidly evolving with evidence of positive selection (Schueler et al., 2010). The N-terminal tail of CENP-A in primates shows evidence of positive selection, including known phosphorylation sites that differ between species.

Despite its extensive functional conservation among eukaryotes, there are a number of species where CENP-A does not function, or is completely absent, at the interface between chromosomes and kinetochores. During meiosis II in *C. elegans* the levels of CENP-A^{HCP-3} associated with chromosomes were shown to be extremely reduced and chromosome segregation does not require CENP-A, suggesting that a CENP-A independent pathway exists for kinetochore function (Monen et al., 2005). Surprisingly, CENP-A has been shown to be absent in four independently derived holocentric groups of insects (Drinnenberg et al., 2014). These sequence analyses also indicated a varying degree of loss of genes

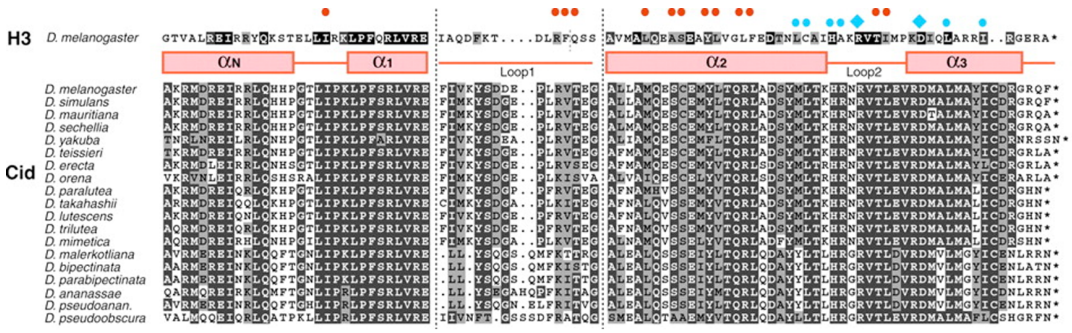


Figure 1.7. Sequence alignment of *Drosophila* CENP-A^{Cid} orthologs (Malik et al., 2002). Sequence alignments are shown for the conserved histone fold domain of the CENP-A^{Cid} orthologs in different *Drosophila* species. Gray and black boxes indicate similar and identical amino acids, respectively. Blue diamonds, blue dots, and red diamonds indicated sites of interaction between H3 and itself, another H3, and H4.

encoding many inner-kinetochore proteins, but the outer-kinetochore components homologous to Ndc80 complex subunits were detected in most species. It was proposed that an ancestral mutation allowed CENP-A to be functionally replaced in these insect groups since all other analysed holocentric species retain a gene that encodes CENP-A. However, these studies could not determine whether the loss of CENP-A triggered the transition to holocentricity or whether CENP-A was rapidly lost following the formation of holocentric chromosomes.

It has been noted that the discussion of extensive conservation of CENP-A (as well as of proteins in general) has focused largely, if not entirely, on model organisms belonging to a single eukaryotic supergroup known as the Opisthokonta (Akiyoshi and Gull, 2013). Kinetoplastids have been put forward as a more evolutionarily diverged eukaryotic species to be developed as a model organism in order to study chromosome segregation. Kinetoplastids include *Trypanosomatida* and *Bodonida*, and split from the main eukaryotic lineage at an early point in evolution. The centromere regions have been identified in *Trypanosoma cruzi* and *Trypanosoma brucei* genomes and based on synteny are conserved in position (Echeverry et al., 2012; Obado et al., 2007). In both species the centromeres formed over regions comprised of retroelement arrays, though the *T. cruzi* centromeres were GC-rich and the *T. brucei* centromeres were AT-rich. *T. brucei* was identified as of interest for the study of chromosome segregation because the genome assembly failed to identify proteins homologous to most of the key conserved centromere and kinetochore proteins, including CENP-A, CENP-C, CENP-E, CENP-F, and Ndc80 (Berriman, 2005). More recently, screening has allowed the identification of nineteen *T. brucei* kinetochore proteins which share no detectable homology with the characterised kinetochore proteins of fungi, plants, or metazoa (Akiyoshi and Gull, 2014). There are six eukaryotic supergroups and the unexpected finding that a distinct set of kinetochore proteins mediate chromosome segregation in a second eukaryotic supergroup completely opens up the question of how exactly conserved either of the two characterised segregation systems are and which, if either, is ancestral.

Centromere DNA evolution

The centromeres sequences are also rapidly evolving, even between closely related lineages. The dimorphic yeasts *C. albicans* and *C. dubliniensis* diverged 20 million

years ago and, although the location of their centromeres has been conserved and the surrounding genes are syntenic, their centromeres do not share any detectable sequence homology (Padmanabhan et al., 2008). Similarly, species in the genus *Schizosaccharomyces* also have conserved centromere locations that are syntenic with flanking genes, but their centromeres exhibit no detectable sequence homology between species (Rhind et al., 2011). Although budding yeast centromeres are rapidly evolving they do not show evidence of selective sweeps, indicating that their rapid evolution is not due to positive selection (Bensasson, 2011; Bensasson et al., 2008). Nevertheless, discovery of sequence dependent point centromeres in the budding yeast *Naumovozyma castellii* that distinct CDE motifs revealed that point centromeres can evolve rapidly (Kobayashi et al., 2015). The rapid evolution of both centromere sequences and CENP-A protein may be linked as it has been proposed that meiotic drive in species with asymmetric meiosis may lead to positive selection of CENP-A that prevents centromeric sequences from expanding as selfish elements (Dawe and Henikoff, 2006; Malik, 2002).

CENP-A nucleosome structure

Canonical histones form an H2A/H2B/H3/H4 octameric nucleosome composed of a tetrameric $[H3\text{-}H4]_2$ core bound by two [H2A-H2B] dimers that wraps ~147 bp of DNA (Luger et al., 1997). Human nucleosomes containing the H3 variant H3.3, has a very similar crystal structure to that of canonical H3 (H3.1) containing nucleosomes (Tachiwana et al., 2011a). The crystal structure of *in vitro* reconstituted CENP-A containing nucleosomes assembled from recombinant proteins expressed in *E. coli* has also been determined (Tachiwana et al., 2011b). The overall structure of these CENP-A nucleosomes is similar to that of canonical H3 nucleosomes (Figure 1.8; Dunleavy et al., 2013). These CENP-A nucleosome are composed of an octameric complex that, despite not being more compact than H3 nucleosomes, only wrap ~121 bp of DNA with partly unwrapped nuclease sensitive DNA ends flanking the core particle (Tachiwana et al., 2011b; Yoda et al., 2000). The protection of less DNA is explained by the shorter alpha-N helix in CENP-A nucleosomes that protects a longer stretch of DNA as it enters and exits in H3 nucleosomes (Tachiwana et al., 2011b; Yoda et al., 2000). This unbound DNA may normally contain the binding sites for the CENP-B DNA binding domain since CENP-B boxes are preferentially located at nucleosome boundaries (Tachiwana et al., 2012). Interestingly, two size classes of CENP-A nucleosomes were detected in

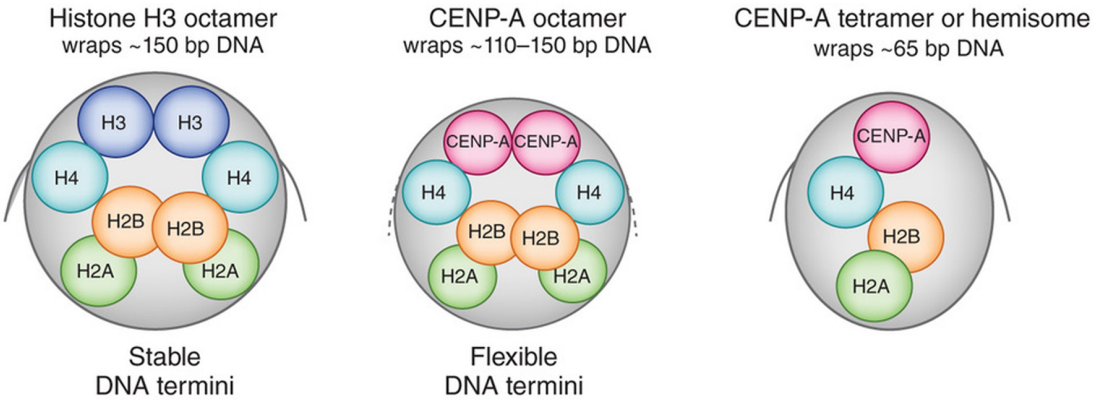


Figure 1.8. Structure of H3 and CENP-A nucleosomes (Dunleavy et al., 2013). Canonical nucleosomes are octameric complexes containing two copies each of histones H2A, H2B, H3, and H4. Canonical nucleosomes wrap ~150 bp of DNA. The centromere specific CENP-A containing nucleosome has a similar structure. It is most likely an octameric complex containing two copies each of H2A, H2B, CENP-A, and H4. CENP-A nucleosome wrap less DNA and have partially unwrapped DNA ends that make it more sensisitive to digestion by MNase. It has also been proposed that CENP-A forms a tetrameric complex containing a single copy of H2A, H2B, CENP-A, and H4, possibly with octameric and tetrameric complexes being present at different stages of the cell cycle. Current evidence favors an octameric CENP-A nucleosome.

CENP-A over-expressing human cancer cell lines where CENP-A is ectopically incorporated. The smaller and larger sizes corresponded to nucleosome wrapping previously described for CENP-A (121 bp) and H3 (147 bp) nucleosomes, respectively (Lacoste et al., 2014). The larger footprint size was only observed for ectopic heterotypic nucleosomes that presumably contained a CENP-A-H3.3 dimer and was correlated with a higher GC-content. Thus the DNA sequence context may also influence the length of DNA wrapped around nucleosomes.

Despite these analyses, the structure of CENP-A nucleosomes remains disputed, with tetrameric and octameric complexes being proposed. Both *in vitro* and *in vivo* MNase digestion of human CENP-A nucleosomes leads to a size distribution consistent with a ~100 bp MNase resistant core with loosely wrapped DNA termini, which is too large for protection by a tetrameric complex (Hasson et al., 2013). Nevertheless, a tetrameric 'hemisome' complex has also been proposed (Figure 1.6; Dalal et al., 2007). AFM analyses indicated that CENP-A nucleosomes isolated from *Drosophila* and human cells are half the height of canonical nucleosomes (Dalal et al., 2007; Dimitriadis et al., 2010) and an CENP-A/H4 tetrameric complex was crystalized in association with the CENP-A chaperone HJURP, which is homologous to Scm3 (Hu et al., 2011). In addition, it has also been proposed that the structure of human CENP-A nucleosomes may differ between cell cycle stages, with both tetrameric and octameric complexes being detected (Bui et al., 2012). Related to this, fluorescence microscopy and FRET analyses of *S. cerevisiae* CENP-A^{Cse4} suggest that the number of CENP-A^{Cse4} molecules per *S. cerevisiae* centromeres varies between one and two across the cell cycle (Shivaraju et al., 2012). Moreover, a hexameric complex consisting of two subunits each of CENP-A, H4, and Scm3, but lacking H2A and H2b, has also been proposed to exist at *S. cerevisiae* centromeres (Camahort et al., 2009; Mizuguchi et al., 2007; Xiao et al., 2011). However, centromeric Scm3 was shown highly dynamic in contrast to the stable centromere association of CENP-A^{Cse4} and it was therefore concluded that Scm3 is likely to be an assembly factor for CENP-A^{Cse4} (Wisniewski et al., 2014).

Other analyses have provided further support for an octameric CENP-A nucleosome model. First, a nanotrap was designed to visualize individual chromatin bound CENP-A complexes with a fluorescent tag isolated from HeLa cells (Padeganeh et al., 2013). Successive photobleaching allowed detection of stoichiometry of CENP-A

in the nucleosomes. The majority of nucleosomes across the cell cycle contained two CENP-A molecules and histone H4 but not HJURP, while a small subset contained a single CENP-A molecule as well as HJURP. The latter species may represent a subset of partially assembled nucleosomes that are in the process of being deposited. Second, CENP-A^{CID} in *Drosophila* S2 cells is isolated as an octameric complex that is less stable than a canonical H3 nucleosomes (Zhang et al., 2012). Crosslinking analyses suggested that the structural basis of dimerization was shared with that of H3 nucleosomes (Zhang et al., 2012). Third, budding yeast CENP-A^{Cse4} containing an internal photobleachable fluorophore showed no evidence of differences in nucleosome structure across the cell cycle (Wisniewski et al., 2014). Reported changes in the number of CENP-A^{Cse4} molecules at centromeres across the cell cycle were attributed to fluorophore maturation and two CENP-A^{Cse4} molecules were detected at each centromere (Wisniewski et al., 2014). Fourth, the most direct evidence comes from AFM studies using recombinant fission yeast and human CENP-A nucleosomes (Miell et al., 2013). Octameric nucleosomes assembled *in vitro* showed a similar height difference depending on whether they contained CENP-A^{Cnp1}, CENP-A or histone H3, and the difference could be attributed to the CENP-A targeting domain (Miell et al., 2013). It was concluded that the difference in height results from CENP-A nucleosomes having a more compact structure.

CENP-A domain function

Human CENP-A is a histone H3 paralog with C-terminal homology to histone H3, but a divergent N-terminal region (Sullivan et al., 1994). The CENP-A histone fold domain is necessary and sufficient for targeting CENP-A to centromeres. Indeed, deletion of the N-terminal tail from CENP-A does not affect the localization of CENP-A to centromeres (Shelby et al., 1997). This function is conserved as the histone fold domain of *Drosophila* CENP-A^{CID} alone can localize CENP-A^{CID} to centromeres (Vermaak et al., 2002). In budding yeast overexpression of the CENP-A^{Cse4} histone fold domain alone was able to confer functionality and CENP-A^{Cse4} also contains an non-essential N-terminal domain (END) (Chen et al., 2000; Morey et al., 2004).

The region of the CENP-A histone fold domain that is sufficient for centromere localisation has been termed the CENP-A targeting domain (CATD) (Sullivan et al., 1994). Expression of a chimeric histone H3 containing the CATD in place of the

normal H3 histone fold domain in HeLa cells allows localisation to centromeres, rescues depletion of endogenous CENP-A, and confers proper chromosome segregation and recruitment of CENP-B and CENP-C (Black et al., 2007a). These functions of the CATD are dependent on the histone fold loop1 and α2 helix and are achieved by interaction of the CATD with the CENP-A chaperone HJURP (Black et al., 2007a; Shuaib et al., 2010). CATD function requires the HJURP mid-domain, which contains a DNA binding region necessary for CENP-A deposition but not localization (Müller et al., 2014). The CATD provide nucleosomes formed by the chimeric $H3^{CATD}$ protein the same increased rigidity shown by CENP-A nucleosomes (Black et al., 2004, 2007b). $H3^{CATD}$ can be propagated at human centromeres following CENP-A deletion, but $H3^{CATD}$ is unable to stably recruit kinetochore components (Fachinetti et al., 2013). The similar CATD of budding yeast CENP-A^{Cse4} (loop 1 and α2 helix) is also sufficient to confer centromere localization (Black et al., 2007a). The centromere specific localization of budding yeast CENP-A^{Cse4} is partly achieved by degradation of non-centromeric, euchromatin associated CENP-A^{Cse4} by the E3 ubiquitin ligase Psh1 and this prevents the ectopic incorporation of CENP-A^{Cse4} (Hewawasam et al., 2010; Ranjitkar et al., 2010). This process depends on specific interactions between CENP-A^{Cse4} and Psh1 mediated by the CATD (Ranjitkar et al., 2010) and is counteracted by association of Scm3 with CENP-A^{Cse4} (Hewawasam et al., 2010).

CENP-A nucleosome numbers and dynamics

Estimates of the number of CENP-A nucleosomes per centromere, even in the same species, vary widely. In fission yeast multiple CENP-A^{Cnp1} nucleosomes assemble on the central domain of each centromere (Appelgren et al., 2003; Coffman et al., 2011; Joglekar et al., 2008; Lawrimore et al., 2011) and estimates using super resolution microscopy suggest that ~12-14 CENP-A^{Cnp1} nucleosomes occupying ~20 possible sites per centromere (Lando et al., 2012). In budding yeast a single CENP-A^{Cse4} nucleosome is thought to occupy each centromere (Furuyama and Biggins, 2007; Henikoff and Henikoff, 2012). At mammalian and *Drosophila* centromeres arrays of CENP-A and H3 nucleosomes are interspersed on repetitive elements such as satellite repeats (Blower et al., 2002; Lam et al., 2006; Vafa and Sullivan, 1997). Recent 3D fluorescent microscopy analyses resulted in estimate of 400 CENP-A molecules per centromere of HeLa cell chromosomes (Bodor et al., 2014). However, this number was found to vary directly with the amount of CENP-A

present within each cell. Differences in the amount of centromeric CENP-A did not affect kinetochore attachments although when the level of CENP-A was reduced below 50% of normal chromosome segregation defects were indicated by the formation of micro-nuclei.

Most H3 nucleosomes are thought to partition randomly during DNA replication and (if this also applies to CENP-A) both chromosomes will receive centromeric CENP-A, making *de novo* establishment unnecessary (Jansen et al., 2007; Xu et al., 2010). Following replication, CENP-A levels at human centromeres drop by half and thus CENP-A levels must be replenished at some point during the cell cycle (Jansen et al., 2007). However, in budding yeast both copies of centromeric CENP-A^{Cse4} are replaced during S phase following centromere replication (Wisniewski et al., 2014). Initial analyses in fission yeast indicated that CENP-A^{Cnp1} may be loaded at two stages of the cell cycle with initial loading during S phase and further loading occurring progressively during G2 (Takahashi et al., 2005; Takayama et al., 2008). However, the counting of CENP-A^{Cnp1} tagged with a photo-activated fluorophore (mEos2) in single cells using super resolution microscopy revealed that CENP-A^{Cnp1} loading is restricted to G2 (Lando et al., 2012). Thus in fission yeast centromeric CENP-A^{Cnp1} levels are at their maximum prior to and during mitosis.

The timing of the deposition of newly synthesised CENP-A is not conserved in other species. In HeLa cells CENP-A synthesis was initially shown to occur in G2 or mitosis, uncoupled from DNA replication, and incorporation of CENP-A into centromeric chromatin was inferred to also occur during this period (Shelby et al., 2000). More recently, pulse labelling of SNAP-tagged CENP-A was used to determine that incorporation occurs only during a short period of the cell cycle between telophase and early G1 (Jansen et al., 2007). Thus, in contrast to fission yeast, human cells undergo mitosis with only ~50% of the maximal level of centromeric CENP-A. Inhibition of Cyclin Dependent Kinase (CDK) activity leads to CENP-A incorporation at human centromeres during G2 and in DT40 cells double mutation of both CDK1 and CDK2 leads to new CENP-A incorporation at centromeres during either S phase or G2, while normal incorporation takes place upon mitotic exit in G1 (Silva et al., 2012). In *Xenopus* CENP-A incorporation occurs during G1 in a manner that is dependent on HJURP (Bernad et al., 2011).

CENP-A requires its chaperone, HJURP in mammals and Scm3 in budding and fission yeast, for incorporation at centromeres (Barnhart et al., 2011; Dunleavy et al., 2009; Foltz et al., 2009; Hayashi et al., 2004; Pidoux et al., 2009; Sanchez-Pulido et al., 2009; Shuaib et al., 2010). The CENP-A^{Cid} chaperone in Drosophila is Cal1, which has no homology to HJURP or Scm3 (Chen et al., 2014). In mammals the recruitment of CENP-A and HJURP to centromeres requires the Mis18α/β and Mis18BP1 proteins (Fujita et al., 2007). Mis18BP1 is recruited to centromeres by CENP-C, thus allowing a reinforcing loop where resident CENP-A recruits new CENP-A via CENP-C→Mis18BP→Mis18α/β→HJURP (Dambacher et al., 2012; Moree et al., 2011). In fission yeast CENP-A^{Cnp1} recruitment also depends on Scm3 and Mis18, as well as Mis16 (Hayashi et al., 2004; Pidoux et al., 2009). Analyses performed in this thesis supported a role for a novel protein, Eic1, in recruiting Mis16 and Mis18 to centromeres, leading to recruitment of Scm3 and CENP-A^{Cnp1} incorporation (Figure 1.9; Subramanian et al., 2014). Thus the role of Eic1 is similar to that of mammalian Mis18BP1 (Subramanian et al., 2014).

CENP-A is a widely conserved centromere specific histone variant and in many species acts as the epigenetic mark for centromere identity. As seen above, centromere organisation varies widely between different species, but they may share the common feature of transcription of centromeric DNA.

1.4 Transcription of centromere DNA

Transcription has been observed to be associated with centromere sequences. Transcription of the outer repeats at fission yeast centromeres has been well characterised. Initially transcripts from centromeres were not detected in wild type fission yeast (Fishel et al., 1988), but transcripts and siRNAs corresponding to the centromeric outer-repeats were subsequently detected (Reinhart and Bartel, 2002; Volpe et al., 2002). Outer-repeat siRNAs are generated from two conserved non-coding regions (Djupedal et al., 2009) that are transcribed by RNA polymerase II (RNAPII) (Djupedal et al., 2005; Kato et al., 2005). These outer repeat transcripts are not easily detectable in wild-type cells but accumulate following deletion of components of the RNA interference (RNAi) pathway such as Dicer (Dcr1), RNA-directed RNA polymerase (RdRP; Rdp1), and Argonaute (Ago1) (Volpe et al., 2002). The reverse strand of outer-repeats are primarily transcribed with much lower levels of transcription of the forward strand (Volpe et al., 2002). Forward strand

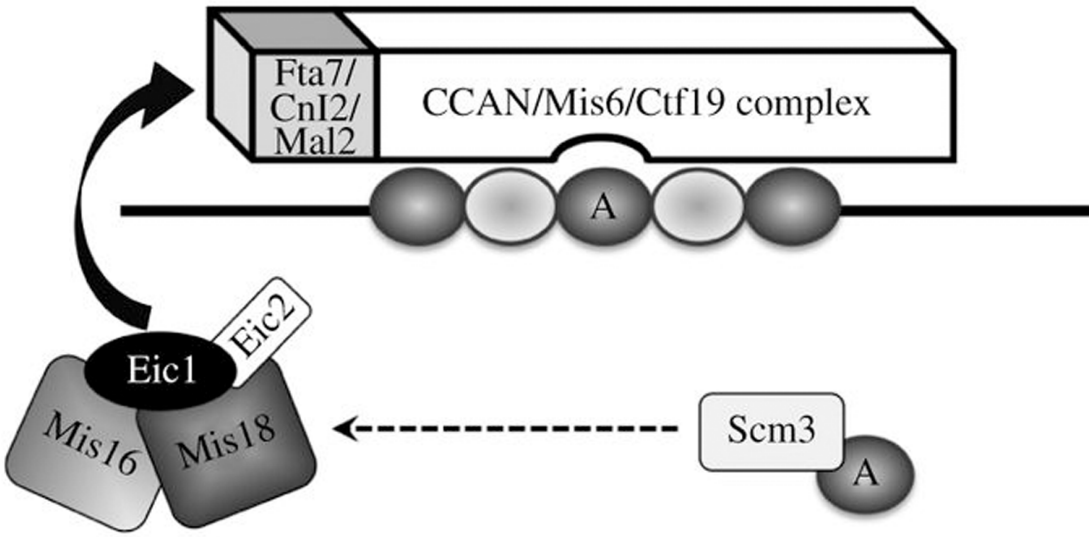


Figure 1.9. A model for Eic1 as a CENP-A^{Cnp1} recruitment factor (Subramanian et al, 2014). Eic1 is recruited to centromeres by components of the constitutive centromere associated network (CCAN). Eic1 interacts with Mis16 and Mis18 to recruit these factors to centromeres. This leads to the recruitment of the CENP-A^{Cnp1} chaperone Scm3 and ultimately to CENP-A^{Cnp1} incorporation.

transcription (which originates at origins of replication within the outer repeats) occurs primarily during S-phase when heterochromatin is partially disrupted and RNAPII occupancy peaks on the outer-repeats, leading to a pulse of siRNA synthesis (Chen et al., 2008; Kloc et al., 2008).

Transcription of the fission yeast centromeric outer-repeats plays a role in heterochromatin structure (Figure 1.10; Holoch and Moazed, 2015). The RNAi machinery is necessary for silencing of transgenes inserted into the outer-repeats and deletion of RNAi components leads to disruption of heterochromatin, including decreased H3K9 methylation (H3K9me) and binding of the HP1 homolog Swi6 (Volpe et al., 2002). The RITS (RNA-induced initiation of transcriptional gene silencing) complex is composed of Ago1, the chromodomain protein Chp1, and a GW domain protein Tas3 (Verdel et al., 2004). siRNAs are loaded into Ago1, allowing RITS to associate with centromeric outer repeats in an RNAi dependent manner (Verdel et al., 2004). Disrupting the interaction between Chp1 and Tas3 leads to loss of RITS localisation to centromeric outer-repeats, loss of centromeric siRNAs and a corresponding accumulation of outer repeat transcripts, and reduction of H3K9me levels at centromeres (Debeauchamp et al., 2008). Tethering Tas3 to a non-centromeric transcript produced from an ectopic locus leads to siRNA generation and RNAi dependent silencing (Bühler et al., 2006). RITS association leads to the recruitment of the RNA-Directed RNA Polymerase Complex (RDRC), containing Rdp1, Hrr1, and Cid12, which synthesizes double stranded RNA (dsRNA) from the nascent RNA template and is required for siRNA generation (Halic and Moazed, 2010; Motamedi et al., 2004; Sugiyama et al., 2005). This dsRNA is cleaved by Dcr1 and incorporated into the RITS complex for further recruitment. The engagement of transcript by Ago1/RITS allows the recruitment of the CLRC complex which contains the H3K9 methyltransferase Clr4 (Shanker et al., 2010; Zhang et al., 2008a). Tethering of Clr4 allows the formation of synthetic heterochromatin in the absence of RNAi at ectopic loci (Kagansky et al., 2009). The methylation of lysine 9 of histone H3 by Clr4 results in the binding of the chromodomain containing proteins Swi6 and Chp2 (both orthologous to HP1), and Chp1, to outer repeat or ectopically methylated chromatin (Ekwall et al., 1996; Kagansky et al., 2009; Shanker et al., 2010; Zhang et al., 2008a).

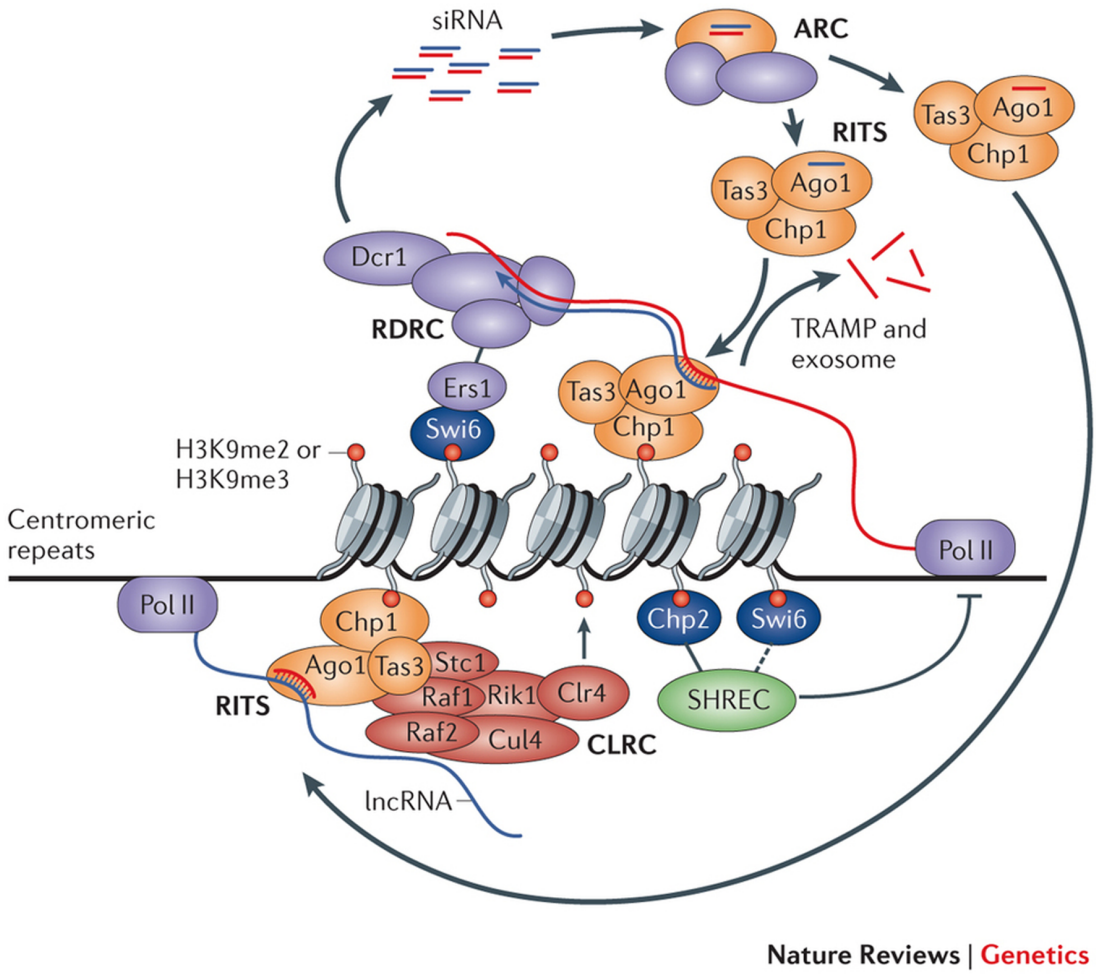


Figure 1.10. RNAi mediated heterochromatin formation in fission yeast (Holoch and Moazed 2015). The RITS (RNA-induced initiation of transcriptional gene silencing) complex associates with siRNA and localises to the centromeric outer-repeats via H3K9me binding activity of Chp1. This leads to recruitment of the RNA-Directed RNA Polymerase Complex (RDRC) complex, containing Rdp1, Hrr1, and Cid12, which synthesises double stranded DNA that is cleaved by Dicer (Dcr1) and incorporated into RITS for further recruitment. This leads to recruitment of the Clr-C complex (CLRC), containing the H3K9 methyltransferase Clr4, which methylates H3K9 and allows binding of the HP1 homologs Chp2 and Swi6.

Transcription also plays a role in forming boundaries between chromatin domains at fission yeast centromeres. Clusters of tRNA genes form barriers that separate the distinct heterochromatin on outer-repeats from the CENP-A chromatin assembled over the central domain where the kinetochore forms (Kuhn et al., 1991; Partridge et al., 2000; Takahashi et al., 1991). Transcription of tRNA genes by RNAPIII is required to maintain the boundary between heterochromatin and CENP-A chromatin and the insertion of a tRNA gene has also been shown to form a boundary that can prevent the spreading of heterochromatin at an ectopic locus (Scott et al., 2006). The boundary activity is not limited to a specific tRNA and is independent of tRNA orientation (Scott et al., 2007).

The central domain of fission yeast centromeres is transcribed by RNAPII, but the function of this transcription is less well understood. These central domain transcripts are rapidly degraded by the exosome (Choi et al., 2011). Promoters within the central domain recruit the chromatin remodeller $Hrp1^{Chd1}$, suggesting that transcription might involve H3 nucleosome eviction (Choi et al., 2011). Other analyses suggest that stalling of RNAPII-mediated transcription of central domain sequences may play a role in CENP-A^{Cnp1} incorporation (Catania et al., 2015).

Centromere sequences in many other species are also transcribed. Maize centromeric CENP-A^{CENH3} occupied CentC repeats are transcribed into long RNA (Du et al., 2010) and Maize CENP-A^{CENH3} has been found to associate with long single-stranded RNAs that originate from the CentC repeats and centromeric retrotransposable elements (Topp et al., 2004). Budding yeast centromere and pericentromere sequences are transcribed by RNAPII (Ohkuni and Kitagawa, 2011). Transcription of the pericentromeric major satellite repeats has been detected in mouse ES cells (Lehnertz et al., 2003) and transcripts from minor satellite repeats, which peak during G2/M and may be degraded or processed into small RNAs that are not detected in other experiments, have also been detected in mouse MEL and NIH3T3 cells (Ferri et al., 2009; Kanellopoulou et al., 2005). Transcription of human centromeric alpha-satellite repeats by RNAPII has been observed during mitosis, which was unexpected as RNAPII is normally removed from metazoan chromosomes during mitosis (Wong et al., 2007). Importantly, active RNAPII was also detected at the non-alpha-satellite containing human mardel(10) neocentromere during mitosis (Chan et al., 2012). Centromeres on human artificial

chromosomes are also transcribed and contain histone modifications that are associated with active transcription (Bergmann et al., 2011).

In contrast, it is also known that high levels of transcription can be incompatible with centromere function. Transcription at budding yeast centromeres impairs centromere function (Ohkuni and Kitagawa, 2011) and transcription from an inducible promoter has been used to regulate centromere activity and create inducible dicentric chromosomes (Mythreye and Bloom, 2003). Similarly, the induction of high levels of centromere transcription on a human artificial chromosome results in defective CENP-A incorporation and chromosome segregation (Bergmann et al., 2012). The neocentromeres induced by centromere deletion in chicken DT40 cells preferentially form over regions containing inactive genes and neocentromere formation was found to be associated with transcriptional inactivation (Shang et al., 2013). However, transcription of genes underlying the human mardel(10) neocentromeres were largely unaffected (Saffery et al., 2003). Fission yeast neocentromeres form over genes that are transcribed at low levels and following neocentromere formation their transcription remains low even during nitrogen starvation when they would normally be up-regulated (Ishii et al., 2008). Thus, proper centromere function may depend on maintaining a sufficiently low level of transcription to aid centromere activity but which does not interfere with kinetochore assembly (Ohkuni and Kitagawa, 2011).

However, the function of transcription at centromeres is not well understood outside of transcription of fission yeast heterochromatin repeats. Transcription of fission yeast central domains may promote the replacement of H3 with CENP-A nucleosomes (Catania et al., 2015; Choi et al., 2011). However, the exact details remain to be determined. In other systems transcription may also be related to CENP-A incorporation. This is supported by the observation that the RNA transcribed from the mardel(10) neocentromere is associated with CENP-A chromatin and its transcription was found to be required to maintain full levels of CENP-A occupancy (Chueh et al., 2009). Furthermore, the transcriptionally active histone modification H3K4me2 is required for CENP-A incorporation and the maintenance of centromere activity on a human artificial chromosome (Bergmann et al., 2011). Alternatively, centromeric transcription could lead to recruitment of other centromeric proteins. RNAPII inhibition during mitosis leads to lagging

chromosomes and decreased levels of CENP-C at human centromeres (Chan et al., 2012). Additionally, human CENP-C has been shown to associate with alpha-satellite RNA *in vitro and in vivo* (Wong et al., 2007) and this association appears to be required for the localisation of CENP-C but not CENP-A to centromeres (Wong et al., 2007). Maize CENP-C has been shown to bind both double-stranded DNA and single-stranded RNA *in vitro* and short single-stranded RNAs stabilise its interaction with DNA (Du et al., 2010).

The potential relationship between centromere transcription and CENP-A incorporation may involve the activity of the FACT (facilitates chromatin transcription) complex and CHD1 (Figure 1.11; Allshire and Karpen, 2008). In chicken DT40 cells components of the FACT complex, which is involved in the transcription coupled disassembly and reassembly of nucleosomes, were found to associate with CENP-A at centromeres, where it recruits the chromatin remodeller CHD1 (Okada et al., 2009). It was proposed that transcription-coupled remodelling activity of $CHD1^{Hrp1}$ within the central domain might play a role in evicting H3 nucleosomes and incorporating of CENP-A^{Cnp1} nucleosomes in their place (Choi et al., 2011). Both FACT and CHD1 were found to be required for proper CENP-A incorporation (Okada et al., 2009). FACT was also found to associate with CENP-A extracted from human cells (Foltz et al., 2006b; Obuse et al., 2004) and CHD1 interacts with FACT components in *Drosophila*, humans, and budding yeast (Kelley et al., 1999; Simic et al., 2003). However, it remains to be determined exactly how FACT and CHD1 influence CENP-A^{Cnp1} incorporation. In fission yeast overexpressing CENP-A^{Cnp1} and loss of FACT function allowed the widespread promiscuous ectopic incorporation of CENP-A^{Cnp1} and prevented its accumulation at subtelomeric regions and non-centromeric central domain sequences (Choi et al., 2012). Loss of FACT allowed replacement of H3 with CENP-A^{Cnp1} on transcribed genes throughout the genome, and consequently CENP-A^{Cnp1} could not accumulate at those regions as normally observed in wild-type cells (Choi et al., 2012). Thus, depletion of FACT may result in CENP-A being redistributed from centromeres and may not be directly involved in CENP-A deposition. However, this does not explain the role of the conserved association observed between FACT and CENP-A. Fission yeast $Chd1^{Hrp1}$ is required for silencing of the outer-repeats and central domain, as well as being necessary for maintaining normal CENP-A^{Cnp1} levels (Walfridsson et al., 2005). A subset of promoters occupied by $Chd1^{Hrp1}$ require

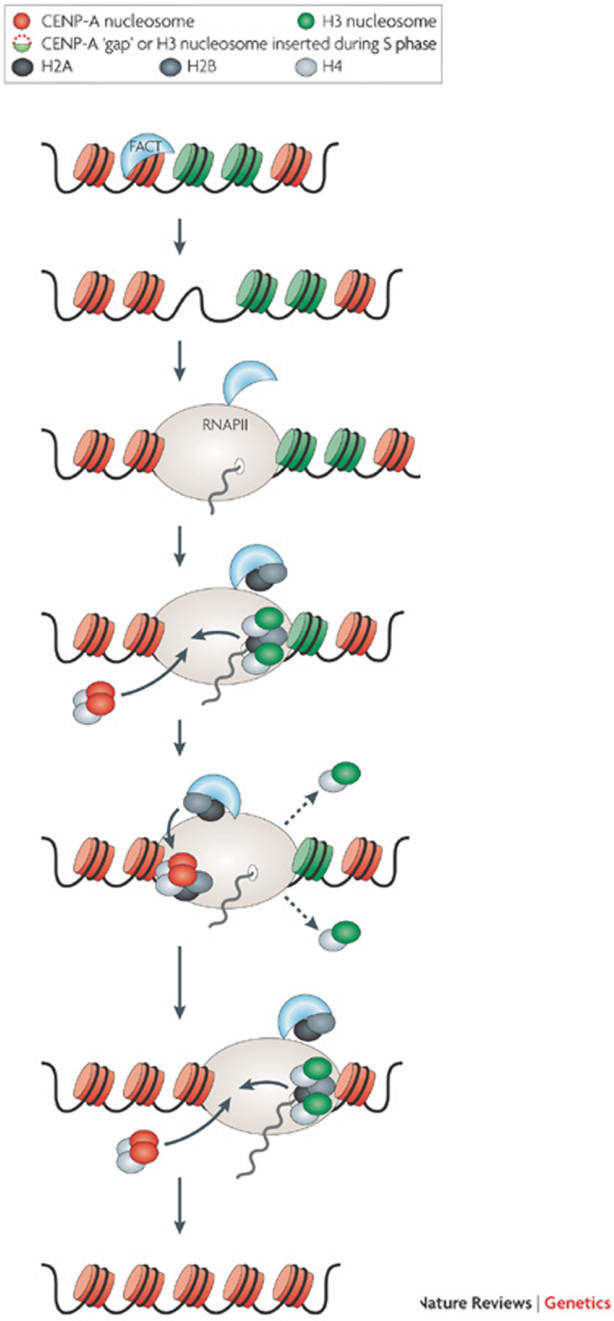


Figure 1.11. Potential role of transcription at the centromere (Allshire and Karpen, 2008). The FACT (facilitates chromatin transcription) complex plays a role in RNAPII transcription and may evict nucleosomes ahead of RNAPII and allow nucleosomes to be incorporated following passage of RNAPII. Transcription of the central domain could lead to FACT mediated H3 nucleosome eviction (green) and replacement by CENP-A^{Cnp1} (red). Chromatin remodelers such as Chd1 may also play a role in this process.

Chd1^{Hrp1} to maintain a nucleosome depleted state (Walfridsson et al., 2005). However, low levels of CENP-A were detectable at these promoters. Chd^{Hrp1} is also required to maintain normal CENP-A^{Cnp1} levels within the central domains and in its absence CENP-A levels decline and H3 levels increase.

It is also possible that, as at the outer repeats in fission yeast, the transcription of centromeric sequences in other species is instead required for RNAi-mediated transcriptional inactivation and heterochromatin integrity. Several studies support such a role for centromere DNA transcription, but this remains to be verified. Similar to fission yeast heterochromatic outer repeats, mouse ES cells lacking Dicer have reduced viability and display increased levels of transcripts from the pericentromeric major satellite repeats, but H3K9me was not found to be affected on these repeats (Murchison et al., 2005). In contrast, another study of Dicer deficient mouse ES cells found elevated transcript levels from both pericentromeric major satellite repeats and centromeric minor satellite repeats along with an overall reduction in H3K9me on these repeats (Kanellopoulou et al., 2005). Moreover, Dicer deficient chicken cells containing a human chromosome display mitotic defects, accumulation of human pericentromeric satellite transcripts, and loss of HP1 localisation to centromeres (Fukagawa et al., 2004). The localisation of mouse HP1α to pericentromeric regions has also been shown to dependent on the binding of its hinge domain to RNA (Muchardt et al., 2002). In *Drosophila* RNAi components have also been reported to be required for the proper localisation of heterochromatic proteins, with mutations leading to mislocalisation of HP1 and H3K9me (Pal-Bhadra et al., 2004). Transcription at centromeres may play a role in heterochromatin structure in these species. Heterochromatin itself may also play a role in the recruitment of centromere or kinetochore proteins, since it is required for establishment of fission yeast centromeres, but this function remains less well understood.

1.5 Nucleosome positioning and occupancy

Nucleosomes are not assembled at all positions in the genome with equal probability and their localisation can be defined in terms of positioning and occupancy (Figure 1.12; Struhl and Segal, 2013). Nucleosome positioning is a measure of whether a nucleosome occupying a particular site always has its midpoint positioned at the same base. The degree of positioning ranges from

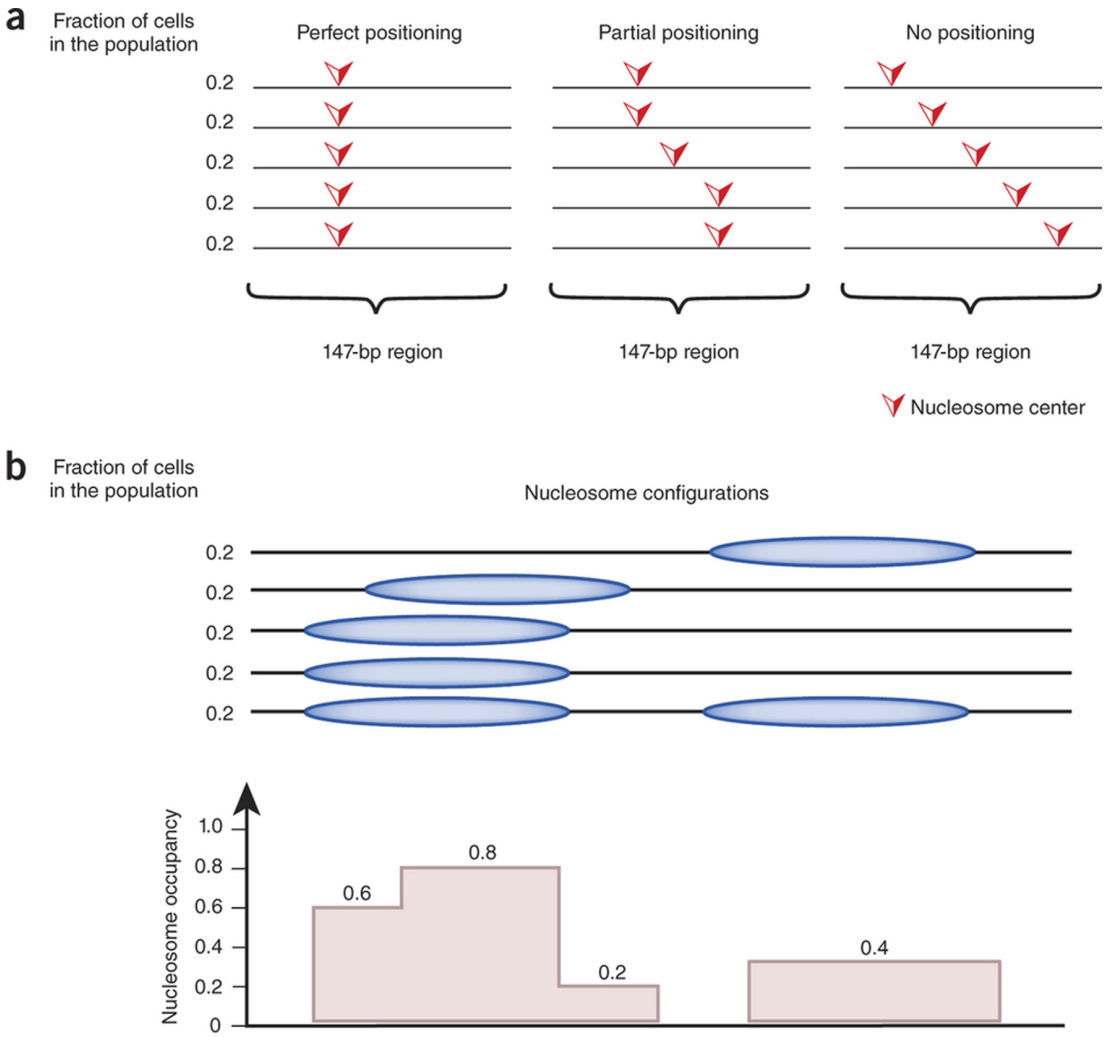


Figure 1.12. Positioning and occupancy of nucleosomes (Struhl and Segal, 2013). There are two concepts to define when considering where nucleosomes occur on DNA: positioning and occupancy. (a) Nucleosome positioning is a measure of constrain on the position of a nucleosome, measured by its midpoint. A perfectly positioned nucleosome is always present at the same 147 bp with its center occuring at the same site. A partially positioned nucleosome usually is present at the same site, but can also be found at positions shifted slightly so that is is also present at different over-lapping sites. A nucleosome that is not positioned can be present at any site within the region without any preference for a particular site. (b) Nucleosome occupancy is a related idea that measures how often in the cell population a nucleosome occurs at a given base. A partially positioned nucleosome with high occupancy (left) has an occupancy profile that shows high occupancy but also has lower occupancy at flanking sites. A highly positioned nucleosome with lower occupancy (right) has lower occupancy since it is present in fewer cells but its occupancy is restricted to a single 147 bp site.

perfectly positioned nucleosomes, with their midpoint always at the same site within the 147 bp bound region, to nucleosomes that are not positioned so that their midpoint occurs at any base within a 147 bp sequence with equal frequency. Nucleosome occupancy measures the probability of any given base pair being occupied by a nucleosome in an individual cell within a population of cells. Synthetic DNA fragments and mouse genomic DNA were both shown to display a large range of nucleosome binding affinities in competition assays, but no genomic sequence was found to possess a similar relative affinity as the most highly occupied synthetic sequences (Thåström et al., 1999). Genomic sequences with TA dinucleotides occurring with a 10 bp periodicity had the highest affinity for nucleosomes. Dinucleotide bendability measured by x-ray crystallography of protein-DNA complexes have indicated that certain dinucleotides, such as TA, have a relatively high level of bendability (Olson et al., 1998). Dinucleotides were subsequently assessed to define motifs that correlate with in vivo nucleosome occupancy in the budding yeast, *S. cerevisiae* (Segal et al., 2006). TT/AA/TA motifs with a 10 bp periodicity were again identified, as well as an alternating CG dinucleotide motif with an ~10 bp periodicity. The periodic nature of these dinucleotides probably reflects the rotational positioning of nucleosomes relative to the bendability of the underlying sequences. Higher resolution maps resulting from cleavage of nucleosomal DNA due to a centrally positioned cysteine in histone H4 supported the relevance of TT/AA/TA periodicity with even greater significance and the refined 10.3 bp periodicity indicated that DNA helix stretches slightly to accommodate nucleosome binding (Brogaard et al., 2012).

Long stretches of A and T, poly(dA:dT) tracks, including those at the promoters of many genes, disfavour nucleosome occupancy *in vivo* in budding yeast (Figure 1.13; Field et al., 2008; W. Lee et al., 2007; Yuan et al., 2005). Although long poly(dA:dT) tracks can wrap around nucleosomes *in vitro* (Prunell, 1982; Schieferstein and Thoma, 1996), their wrapping around nucleosomes is disfavoured (Kunkel and Martinson, 1981). Overall, a high level of correlation was evident between the sites occupied by nucleosomes *in vivo* in budding yeast and the sites where chicken nucleosomes assembled on budding yeast genomic DNA *in vitro*, although there were also some differences observed (Kaplan et al., 2009). Notably, both poly(dA:dT) tracks and transcription start sites were found to be depleted for nucleosomes both *in vivo* and *in vitro*, but the paucity of nucleosomes at promoters

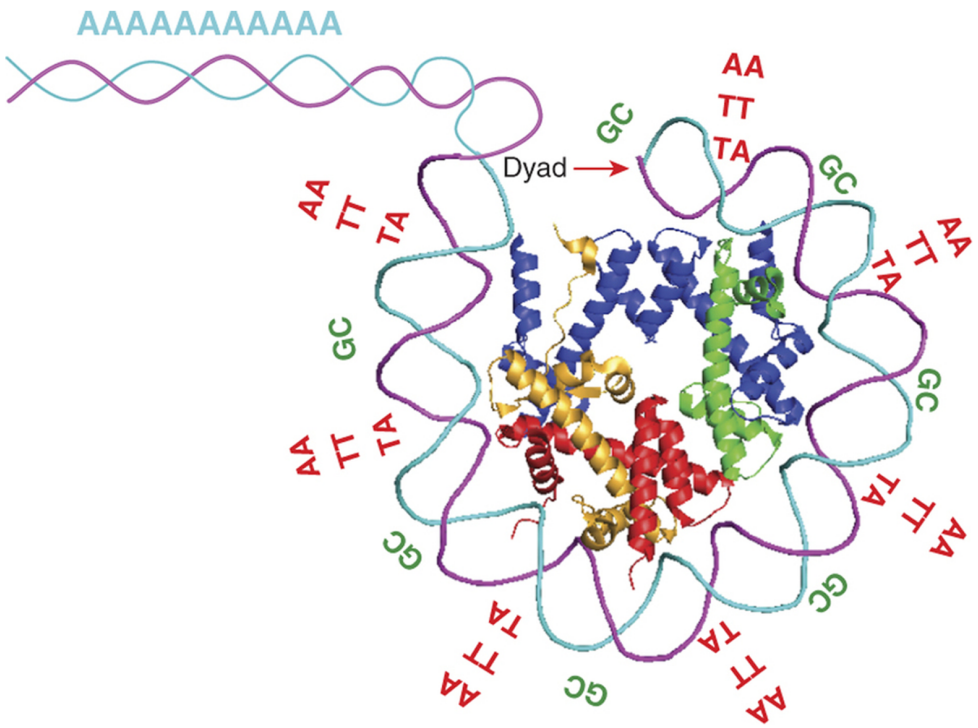


Figure 1.13. Sequence determinants of nucleosome positioning and occupancy (Struhl and Segal, 2013). Long poly(dA:dT) tracks are not preferred sites for nucleosomes to occupy so AT-rich sequences in the genome form nucleosome depleted regions. Positioned nucleosomes show a characteristic sequence pattern with a dinucleotide periodicity that features AA/AT/TT dinucleotides occuring every ~10 bp alternating with CG dinucleotides that occur every ~10 bp.

was more pronounced *in vivo*. Furthermore, some promoter sequences that were depleted for nucleosomes *in vitro* were found to be occupied *in vivo*. It should be noted that although nucleosome occupancy *in vivo* and *in vitro* was generally dependent on similar sequence features, nucleosomes assembled *in vitro* showed less precise positioning (Zhang et al., 2009). Such analyses suggests that, apart from the sequence dependent features identified *in vitro* and *in vivo*, additional *in vivo* factors and biological activities are required for proper formation of nucleosome depleted regions and influence nucleosome occupancy in general.

Many promoters of *S. cerevisiae* genes contain a nucleosome depleted region that is flanked by highly positioned +1 and -1 nucleosomes on either side. Nucleosomes within the 5' end of the gene exhibit precise positioning that gradually decays with distance from the transcription start site so that the +1 nucleosome is highly position but positioning of the +2, +3, +4 becomes sequentially less pronounced (Figure 1.14; Mavrich et al., 2008). The observed ~10 bp TA/AT dinucleotide periodicity was found to be strongest for both the +1 and -1 nucleosomes and consequently may be less relevant for nucleosome positioning residing elsewhere in the genome. The phased nucleosomes at the 5' end of most genes is thought to reflect statistical positioning resulting from the boundary element created by the nucleosome depleted region that constrains the possible position the +1 nucleosome and thus neighbouring nucleosomes (Mavrich et al., 2008). However, this does not explain why nucleosomes downstream of the +1 nucleosome are phased while those beyond the -1 nucleosome show no ordered phasing pattern. The fact that the *in vitro* reconstruction of nucleosomes on genomic DNA does not result in a highly positioned +1 nucleosome and phasing of downstream nucleosomes also indicates that *in vivo* activities such as transcription associated chromatin remodelling are required to generate these distinctive nucleosomal patterns. Consistent with this, these characteristics can be induced on *in vitro* assembled templates the addition of cell extract and ATP (Zhang et al., 2011). Nucleosome depleted regions have been shown to have clear effects on the expression of associated genes. Increasing the poly(dA:dT) track length causes a progressive increase in reporter gene expression levels by reducing the probability of a nucleosome occupying the transcription factor binding site (Raveh-Sadka et al., 2012).

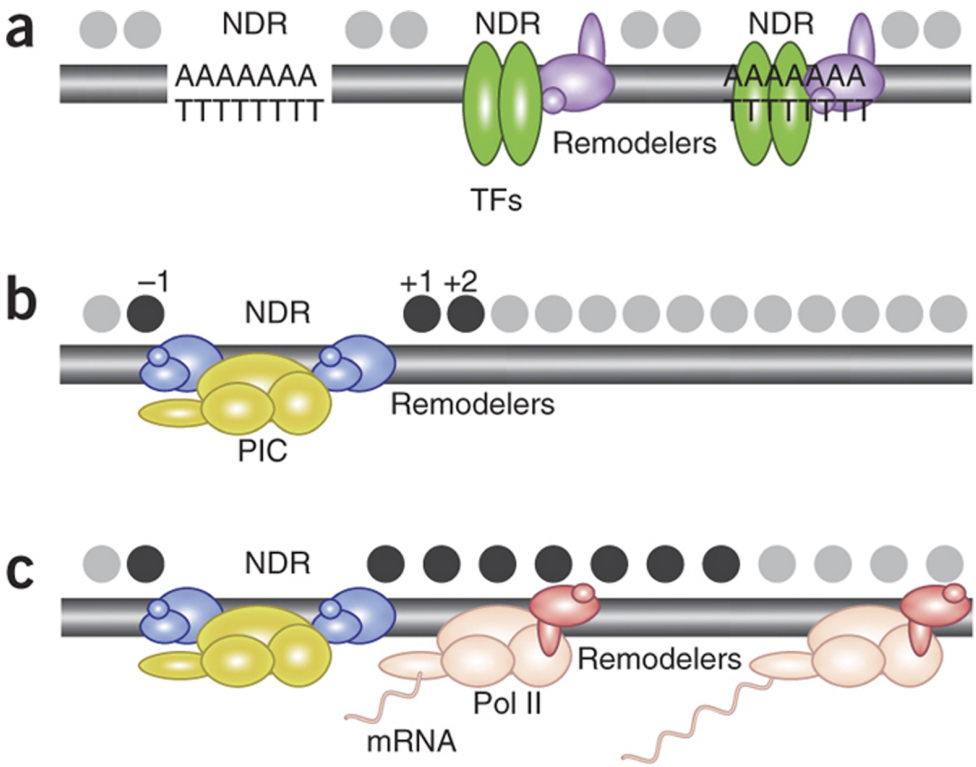


Figure 1.14. Nucleosome depleted regions in the genome (Struhl and Segal, 2013). Multiple mechanisms create nucleosome depleted regions (NDR) and positioned nucleosomes in the genome. Nucleosomes are indicated in grey. (a) Long poly(dA:dT) tracks create nucleosome depleted regions but do not position flanking nucleosomes based on sequence alone. (b) NDRs can also be created by binding of specific factors, such as chromatin moddifiers and the RNAPII pre-initiation complex (PIC). These complexes create boundaries that position flanking nucleosomes. (c) Transcription and recruitment of nucleosome-remodelers also positions nucleosomes downstream.

The extent to which DNA sequence alone influences nucleosome positioning has also been investigated. When sequences from other yeast species (*K. lactis, K. waltii*, and *D. hansenii*) were introduced on artificial chromosomes into S. cerevisiae the nucleosome depleted regions at promoters associated with poly(dA:dT) tracks were largely maintained (Hughes et al., 2012). However, the nucleosome depleted regions associated with promoters that lack poly(dA:dT) tracks were not maintained to the same extent, suggesting that some promoter nucleosome depleted regions are maintained by transcription initiation and chromatin remodelling activities rather than innate DNA sequence features (Hughes et al., 2012). Additional nucleosome depleted regions were also observed within coding regions that happened to contain sequences that defined transcription factor binding sites in *S. cerevisiae*. Moreover, the phased nucleosomes downstream of the +1 nucleosomes of genes exhibited spacing that was indicative of the normal nucleosome spacing observed in the host *S. cerevisiae* rather than that normally observed in the donor species. A comparison of nucleosome positioning in *Hemiascomycota yeasts* also found support for genetically determined nucleosome depleted regions defined by poy(dA:dT) tracks, but also identified nucleosome depleted regions that are dependent on binding sites for chromatin remodelling activities (Tsankov et al., 2010). In contrast, the promoter regions of *S. pombe* (fission yeast) have been shown to be not enriched for poly(dA:dT) tracks and models trained on *S. cerevisiae* and *S. pombe* promoter sequences perform poorly at cross species predictions (Lantermann et al., 2010). In *Ascomycota* fungi poly(dG) tracks have also been shown to be unfavourable for nucleosome occupancy, a pattern that was also observed *in vitro* (Tsankov et al., 2011). The prevalence of poly(dG) and poly(dA:dT) tracks was at least partially related to the nucleotide composition of the genomes of different species analyses.

Transcription may play an additional role in nucleosome positioning *in vivo* since the location of the +1 nucleosome is strongly correlated with the transcription start site of genes (Zhang et al., 2009). When nucleosome positioning was examined on DNA sequences from other yeasts in *S. cerevisiae* (see above) the position of the +1 nucleosome differed from its position in the donor species in a manner that was highly correlated with changes in the positions of transcription start sites (Hughes et al., 2012). Transcriptionally regulated *S. cerevisiae* genes that have multiple levels of transcription also display multiple stable phased configurations with differing

nucleosome densities that potentially reflect changes in transcriptional activity (Vaillant et al., 2010). The 5' ends of genes with variable transcription levels and with multiple nucleosome configurations have also been found to be associated with less H2A.Z, a histone H2A variant. This lack of H2A.Z may reflect the fact that these nucleosomes containing H2A.Z are more stable and highly positioned to phase downstream nucleosomes. Higher transcription rates have also been correlated with more densely phased nucleosome arrays over corresponding genes

As mentioned above, the positioning of CENP-A nucleosomes may be an important aspect of centromere organisation. Therefore, understanding the sequence features, factors and activities that influence nucleosome positioning should provide additional insight into centromere organisation.

1.6 Replication proteins and centromeres

Replication origins

Cells face the challenge of replicating their genetic material once per cell cycle. DNA synthesis is nucleated at replication origins. Hundreds to thousands of replication origins are found throughout eukaryotic genomes. It was initially observed in *S. cerevisiae* that small DNA sequences were able to initiate DNA synthesis on an ectopic plasmid and these were termed autonomously replicated sequences (ARS elements). In agreement with this, *S. cerevisiae* replication origin identity is determined in a partially sequence dependent manner with the 11-bp ARS consensus sequence (ACS) and less conserved B elements being present at all origins of replication (Chang et al., 2011; Eaton et al., 2010; Rehman et al., 2006; Xu et al., 2006). The origin recognition complex (ORC) binds to origins of replication and has subsequently been used to identify origins more directly genome-wide. ORC recruits other replication proteins (see below) to origins in order to license origins to fire during S phase (Figure 1.15; Aladjem, 2007). This results in the synthesis of DNA by bidirectional replication forks, originating from origins, that progress along the chromosome until meeting another fork or a chromosome end. ORC based genome-wide replication origin maps show that *S. cerevisiae* origins are preferentially located in intergenic regions (Wyrick et al., 2001; Xu et al., 2006). ACS elements are present at locations that do not form replication origins and therefore other features must also contribute to origin identity (W. Xu et al., 2006). One possible additional feature is the presence of a nucleosome depleted region, which

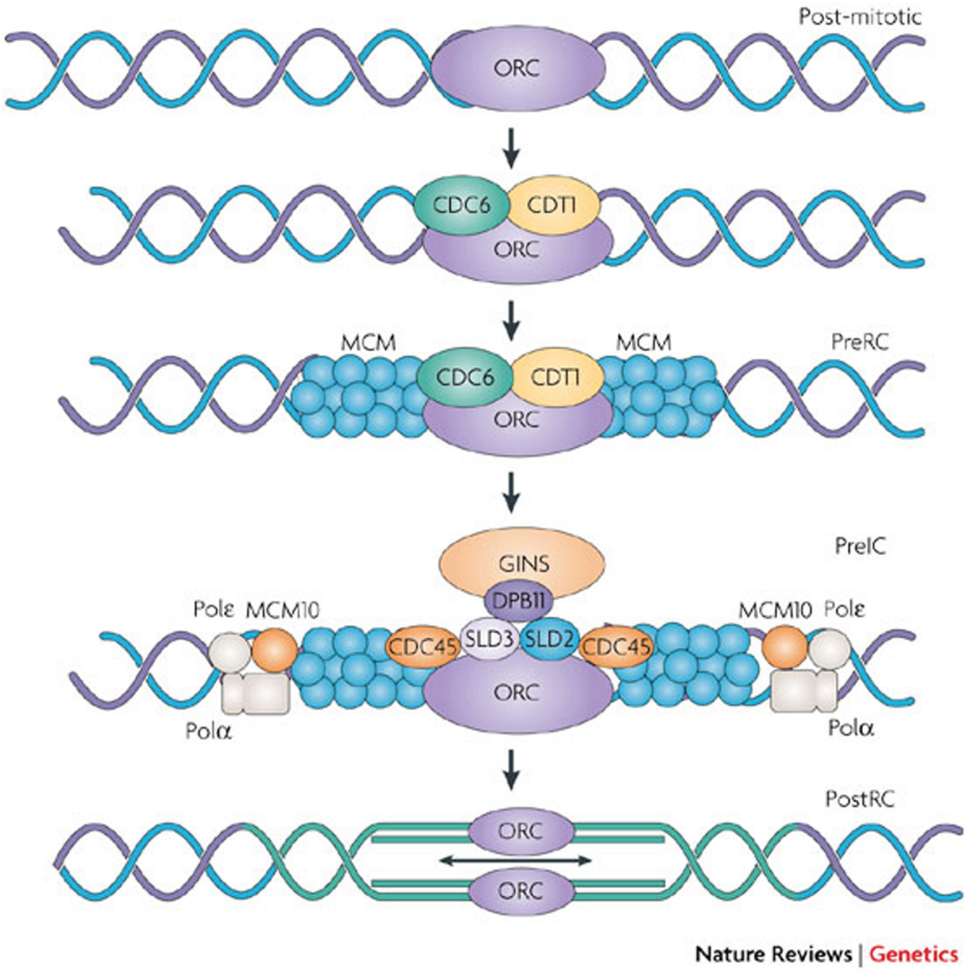


Figure 1.15. The origin recognition complex and replication (Aladjem, 2007). The origin recognition complex (ORC) is a six-subunit complex that binds to replication origins. ORC acts as a platform for recruitment of Cdc6 and Cdt1. Subsequently, the minichromosome maintenance (MCM) proteins, forming a six-subunit helicase complex, are recruited to form a pre-replicative complex (PreRC). Additional factors are recruited to license the origin and create a pre-initiation complex (PreIC) that can initiate DNA synthesis during S-phase. Upon entry into S-phase a subset of origins fire, leading to bidirectional DNA replication from the origin.

is found at known replication origins but not at non-origin ACS elements (Eaton et al., 2010). The nucleosomes that flank active replication origins have been shown to be highly positioned by the binding of ORC and Abf1 (Lipford and Bell, 2001).

As with centromeres, it was assumed that the identification of *S. cerevisiae* replication origins would facilitate the identification of replication origins in other species based on similar sequence features. However, again as with centromeres, identifying sequences that specify replication origins in other species turned out to be far more complicated than just searching for homologous sequences features. It turned out that *S. pombe* replication origins are AT-rich but lack other more specific DNA sequence features (Houchens et al., 2008; Kong and DePamphilis, 2001, 2002; Lee et al., 2001; Segurado et al., 2003). Mammalian replication origins initially evaded stringent characterisation since mammalian replication origins do not exhibit ARS activity on extrachromosomal plasmids (Schaarschmidt et al., 2004). However, the mapping of ORC localisation in genome-wide studies has shed light on the sequence features underlying replication origin identity in mammals as well as *Drosophila*. Such analyses show that origins are associated with AT-rich sequences in both mammals and *Drosophila* (Cadoret et al., 2008; Delgado et al., 1998; Eaton et al., 2011; Gómez and Antequera, 2008; Liu et al., 2012; MacAlpine et al., 2004; Prioleau et al., 2003; Sequeira-Mendes et al., 2009). In support of this, it was found that human ORC preferentially binds to non-specific AT-rich DNA *in vitro* (Vashee et al., 2003). However, further analyses revealed that an asymmetric G-rich motif was the primary sequence feature associated with human and Drosophila replication origin identity (Cayrou et al., 2011). Replication origins are preferentially located near transcribed genes in mammals and *Drosophila* and are associated with CpG islands and CpG-like regions, respectively (Cadoret et al., 2008; Cayrou et al., 2011; Delgado et al., 1998; Gómez and Antequera, 2008; Ladenburger et al., 2002; Sequeira-Mendes et al., 2009). These analyses indicated that replication origin identity is quite distinct in fission yeast and mammals, but it is possible that there are some underlying mechanistic similarities between fission yeast and metazoan replication origins. In mouse cells CG-rich sequences have been shown to exclude nucleosomes and thus nucleosome exclusion may be one important and conserved feature involved in origin identity (Fenouil et al., 2012). Indeed, it has been proposed that the property of nucleosome exclusion, rather than specific sequence features, is

the primary determinant of origin identity and allows ORC binding to origin DNA (Xu et al., 2012).

Replication proteins

In contrast to the varying sequence features found at replication origins, the proteins involved in recognising replication origins are highly conserved among eukaryotes. ORC is a widely conserved six-subunit complex, consisting of Orc1-6, that binds to DNA at replication origins (Dhar and Dutta, 2000; Duncker et al., 2009; Foss et al., 1993; Grallert and Nurse, 1996; Kong et al., 2003; McNairn et al., 2005; Zisimopoulou et al., 1998). ORC binds to DNA in nucleosome depleted regions and generates a nuclease resistant footprint of ~130 bp in *Drosophila* (MacAlpine et al., 2010; Remus et al., 2004). Similarly, AFM measurements of fission yeast ORC showed that it wraps ~140 bp of DNA (Gaczynska et al., 2004). *S. pombe* ORC is recruited to AT-rich DNA due to the nine AT-hooks of Orc4 that preferential bind asymmetric AT-rich DNA stretches (Chuang and Kelly, 1999; Kong and DePamphilis, 2001; Lee et al., 2001). The presence of AT-hooks domains in the Orc4 protein is restricted to all fission yeast species despite the wide conservation of Orc4 itself (Chuang and Kelly, 1999; Lee et al., 2001; Mojardín et al., 2013). The presence of AT-hooks is therefore likely be related to the general AT-rich DNA sequence content found at fission yeast origins. Apart form Orc4, the Orc2 subunit in *S. cerevisiae* and the fission yeast *Schizosaccharomyces japonicus* contains AT-hook domains, although their functional significance has not been explored (J. Xu et al., 2012). None of the human ORC subunits contain AT-hook domains, but human ORC has been shown to interacts with the AT-hook containing protein HMGA1a and may localize to some sequences through this interaction (Thomae et al., 2008, 2011). However, the importance of HMGA1a and its AT-hook domain for recruiting ORC is now unclear given that asymmetric G-rich motifs appear to determine origin identity in mammals and *Drosophila.*

ORC binding defines the location of replication origins, but other proteins must be recruited to license origins in order to initiate DNA synthesis upon entry into S phase. During G1 ORC recruits Cdc6/18 and Cdt1 and subsequently the minichromosome maintenance (MCM) complex is recruited to origins, where it forms a ring structure around the DNA at each side of an origin (Asano et al., 2007; Chen et al., 2007; Donovan et al., 1997; Hua and Newport, 1998; Kawasaki et al., 2006; Kearsey et al.,

2000; Matsunaga et al., 2001; Prokhorova and Blow, 2000; Remus et al., 2009; Takahashi et al., 2003; Takara and Bell, 2011; Tsakraklides and Bell, 2010; Tsuyama et al., 2005; Weinreich et al., 1999). The MCM helicases are part of a highly conserved family of AAA+ ATPases, each possessing an MCM box, and together form a six-subunit helicase complex consisting of Mcm2-7 (Iyer et al.; Neuwald et al., 1999). All components of the MCM complex are required to form licensed replication origins and initiate DNA synthesis (Labib et al., 2000, 2001; Schwacha and Bell, 2001). MCM loading licenses origins for replication, meaning that origins are capable of firing a single time during S phase, and reflects the formation of a pre-replicative complex (Pre-RC). Restricting origin licensing to G1 prevents re-licensing and origin firing and thus restricts origin firing to only once during S phase. MCM is loaded onto chromatin throughout G1 in excess relative to the number of ORC complexes present in *Xenopus* early embryo and higher levels of Cdc45 bound to the MCM complex correlate positively with origin strength (Edwards et al., 2002a). In *Xenopus* and humans MCM is not restricted to origins, but instead MCM can be observed broadly distributed throughout chromatin (Edwards et al., 2002b; Harvey and Newport, 2003; Ritzi, 1998). The reason for this excess loading of MCM proteins is not known, but experiments with *Drosophila* suggest that MCM may be repositioned by transcription in order to form dormant replication origins that may only fire under conditions of replicative stress (Powell et al., 2015). Other processes can also affect origin licensing and ultimately determine what subset of origins fires during S phase. Shifting the positioned flanking nucleosomes away from *S. cerevisiae* replication origins can prevent origin licensing even though ORC remains bound to DNA (Lipford & Bell, 2001). Related to the chromatin context of origins is the finding that following transcription across an origin replication proteins dissociate and it is necessary to reload replication proteins to relicense the origin (Lõoke et al., 2010).

Upon entry into S phase the kinases Cdc7, Cdc2/Cdk2 and the initiation factor Cdc45 are recruited to origins to trigger DNA synthesis (Tanaka and Araki, 2010). Only a subset of origins fire in any given cell cycle and different origins have different probabilities of firing, which may depend on levels of MCM loaded at individual origins (Grishina and Lattes, 2005; Jares and Blow, 2000; Jares et al., 2004; Lei and Tye, 2001; Pasero et al., 1999; Patel et al., 2008; Yang et al., 2010). There are also dormant origins present in the genome that do not fire under normal

conditions but can fire following replicative stress (Courbet et al., 2008; Ge et al., 2007; Schultz et al., 2010; Woodward et al., 2006). To ensure that the genome is only replicated a single time it is necessary to prevent origins from firing again during S phase. Additional rounds of replication in a single S phase are prevented by the accumulation of Cdc2 and the releasing of MCM and Cdc6/18 from origins, thereby preventing them from being relicensed in that S phase (Hua et al., 1997; Kuipers et al., 2011).

Replication Timing

Certain regions in the genome have characteristic timing of DNA replication during S phase. In particular, centromeres and heterochromatic domains generally show distinct replication timing. In *S. cerevisiae* (McCarroll and Fangman, 1988), fission yeast (Hayashi et al., 2009), and *C. albicans* (Koren et al., 2010) replication of centromere DNA occurs early in S phase. In *S. cerevisiae* Sld3-Sld7 loading by DDK is responsible for the early replication of centromeres (Natsume et al., 2013). This timing potentially has functional importance as neocentromeres at ectopic loci in *C. albicans* (Koren et al., 2010) and centromeres located at ectopic locations in budding yeast (Pohl et al., 2012) replicate early, even though these regions would normally replicate later during S phase. Neocentromere formation in *C. albicans* also leads to ORC recruitment, further suggesting that replication timing is functionally important (Koren et al., 2010). Early replication of *S. pombe* centromeres depends on origins present within the heterochromatic outer-repeats that require Swi6 for their very early replication activity (Kim et al., 2003; Li et al., 2011). ORC has also been detected in the *S. pombe* central domains but this region is replicated passively by replication forks originating from the flanking outer repeats (Hayashi et al., 2007; Kim et al., 2003; Segurado et al., 2003). In other species centromeres also display specificity in timing of centromere replication, but the timing itself differs from that of the species discussed. In chickens (Shang et al., 2013), *Drosophila* (Sullivan and Karpen, 2001), and humans (Ten Hagen et al., 1990; Hultdin, 2001) centromeres replicate during mid to late S phase and chicken neocentromeres shift their replication timing to that of canonical centromeres (Shang et al., 2013).

Replication protein function outside of DNA replication

Replication proteins, such as ORC subunits, can also function in processes outside of replication. A well-known example of this is the role that *S. cerevisiae* ORC plays in silencing of the mating type loci. The mechanism of budding yeast heterochromatin formation differs greatly from that of mammals and fission yeast since it lacks RNAi, H3K9 methylation, and HP1 orthologs (Harrison 2009). Silencing of the *S. cerevisiae* cryptic mating type loci (HML ad HMR) and at telomeres is mediated by the silent information regulator (Sir) proteins, Sir1-4 (Figure 1.16A; Hickman et al., 2011). The E and I silencers at the cryptic mating type loci contain ARS elements that provide binding sites for ORC. In this context ORC acts as a platform to recruit the Sir proteins via an interaction between the Orc1 BAH domain and the Sir1 ORC interacting region (Gardner et al., 1999; Hou et al., 2005; Triolo and Sternglanz, 1996; Zhang et al., 2002). Sir1, along with the DNA binding proteins Rap1 and Abf1, recruit Sir2, Sir3 and Sir4 to the silencing elements (Gasser and Cockell, 2001; Moretti and Shore, 2001; Triolo and Sternglanz, 1996). The Sir2 HDAC deacetylates H4K16, thereby allowing the Sir complex to spread along chromatin using the binding of Sir3 to deacetlyated H4 (Rusché et al., 2002; Tanny et al., 1999; Yang et al., 2008). The silencing pathways are partially redundant as silencing can also be achieved by the SUM1 complex and ORC (Sutton et al., 2001). Thus ORC is required to recruit silencing proteins to HML and HMR in *S. cerevisiae* independently of its role in replication.

ORC is also directly involved in forming silent chromatin at the cryptic mating type loci in other yeasts. The phylogenetic branch of the budding yeasts leading to *S. cerevisiae* contains a whole genome duplication and therefore the Sir proteins have paralogs in some budding yeast species (Figure 1.16B; Hickman et al., 2011). Sir1 paralogs have undergone expansions and contractions and the number of Sir1 paralogs present in different related species varies, with no gene encoding Sir1 being present in *C. glabrata and Kluyveromycis lactis* (Gallagher et al., 2009). The Sir1 paralogs, *termed k*in *o*f *S*ir1 (Kos; Kos1 to Kos4), are found in *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, and *S. castellii* (Gallagher et al., 2009). The Kos proteins appear to have a conserved function in silencing since all *S. bayanus* Kos proteins localise to mating type locus silencers (Gallagher et al., 2009). In these species Hst1, a paralog of Sir2, provides the deacetylase function. Hst1 is part of the SUM1 complex that is involved in silencing (Sutton et al., 2001). Intriguingly,

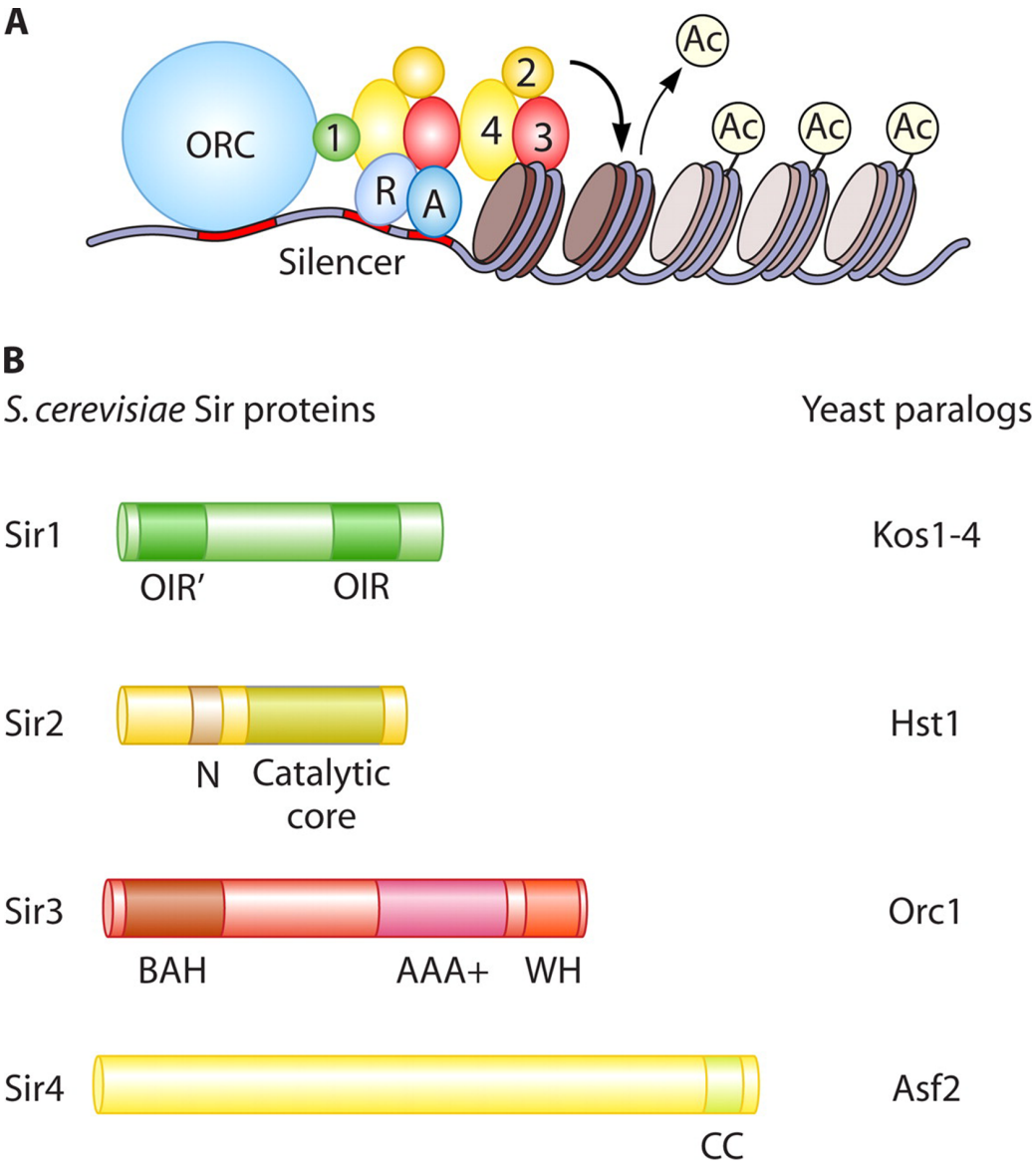


FIgure 1.16. The origin recognition complex functions in budding yeast silencing (Hickman et al., 2011). (A) The origin recognition complex (ORC) binds to DNA along with Rap1 (R) and Abf1 (A). Orc1 interacts with the ORC interacting region (OIR) of Sir1 to anchor the other Sir proteins. Sir2 deacetylates H4K16 to allow Sir3 and Sir4 to bind and the complex to spread. (B) Budding yeast underwent a whole-genome duplication event. Therefore, the Sir proteins also have paralogs in *S. cerevisiae* and other budding yeast species. Sir1 has multiple paralogs, Kos1-4, that have expanded and contracted in related species and act in silencing. The histone deacetylase Hst1 is a paralog of Sir2. Sir3 is a paralog of Orc1. Sir4 is a paralog of the silencing antagonising Asf2.

Orc1 is a paralog of Sir3 and, in fact, Orc1 in these species is more similar to the ancestral Orc1/Sir3 protein (Kellis et al., 2004). Thus, *S. cerevisiae* Sir3 may have been adapted from ancestral Orc1/Sir3 for a specialised role in silencing. Experiments in *K. lactis* support this view since, although it lacks Sir3, it utilises Orc1, along with other Sir proteins, to form silent chromatin at the mating type loci and telomeres (Hickman and Rusche, 2010).

Related to this, ORC also functions in heterochromatin structure in mammals and *Drosophila*. The localisation of ORC and the heterochromatic protein HP1 to heterochromatic foci are inter-dependent in both human and *Drosophila* cells (Auth et al., 2006; Badugu et al., 2005; Pak et al., 1997; Prasanth et al., 2010; Shareef et al., 2001, 2003). ORC is required for heterochromatin integrity and proper sister-chromatin cohesion and defective ORC or HP1 localisation leads to chromosome segregation defects (Auth et al., 2006; Badugu et al., 2005; Pak et al., 1997; Prasanth et al., 2010; Shareef et al., 2001, 2003). The inter-dependence of *Drosophila* ORC and HP1 is due to their direct interaction and both proteins also interact with HOAP (HP1/ORC associated protein) to prevent chromosome fusions and maintain telomere structure (Badugu et al., 2003; Cenci et al., 2003; Pak et al., 1997; Prasanth et al., 2004, 2010; Shareef et al., 2001). Moreover, the mammalian pre-replication complex component known as ORCA (Origin Recognition Complex-Associated) is required to maintain proper levels of H3K9 methylation at heterochromatic loci (Giri et al., 2015). Such observations stress the fact that ORC and associated proteins can influence the state nearby chromatin independently of its role in replication.

The alternative functions of ORC make it an interesting candidate for further study in *S. pombe*. In particular, no known function has been ascribed to the ORC that has been shown to associate with the central domains of fission yeast centromeres. This is particularly intriguing since replication is not initiation from these centromeric ORC bound sites during S phase (Kim et al., 2003; Smith et al., 1995). Since ORC has been shown to influence nucleosome phasing and chromatin structure it is possible that this centromeric ORC plays a role in centromere architecture or determining centromere identity.

1.7 Aims of this thesis

The above discussion summarises the general mechanisms that contribute to chromatin organisation and centromere identity. Some of these mechanisms are not fully understood. This thesis seeks to investigate the evolution and organisation of sequence features within fission yeast centromeres and how they relate to DNA replication, nucleosome positioning, and kinetochore interactions.

Chapter 3: Conservation of sequence features that position CENP-A^{Cnp1} in fission yeasts.

Regional centromeres with epigenetically determined centromere identity are paradoxically associated with particular sequences. Unlike in vertebrates and *Drosophila,* fission yeast CENP-A^{Cnp1} is known to occupy the centromere central domains, which are primarily non-repetitive. The aim of this chapter was to more precisely map the location of CENP-A^{Cnp1} nucleosomes relative to underlying DNA sequences to determine sequence features that may be important for centromere function and identity. CENP-A^{Cnp1} occupied sites in *S. pombe* and *S. octosporus* were mapped at centromeres, as well as at *S. pombe* neocentromeres. The analyses presented show that CENP-A^{Cnp1} occupied sites are highly positioned by large asymmetric AT-rich gaps at centromeres of both species, indicating that this may be a conserved feature of fission yeast centromeres. In contrast, CENP-A^{Cnp1} occupied sites are not strongly positioned by sequence features at neocentromeres, indicating that the AT-rich gaps are a part of optimal centromere organisation but are dispensable for centromere function. In particular, the identified gaps are potential binding sites for the origin recognition complex (ORC).

Chapter 4: Analysis of CENP-A^{Cnp1} distribution following CENP-A^{Cnp1} over-expression.

The analyses presented in chapter 3 reveal that CENP-A^{Cnp1} occupancy at fission yeast centromeres is highly related to underlying sequence features. When CENP-A is overexpressed in fission yeast it is known to accumulate at additional locations. The aim of this chapter was to determine if these additional locations are associated with particular sequence or other features and whether these sequences are related to those identified at centromeres. Therefore, we investigated whether sequence features may play a role in ectopic CENP-A^{Cnp1} incorporation, which may be important for neocentromere formation and CENP-A^{Cnp1} targeting. Chromatin

structure or nuclear organisation likely plays a much larger role in determining sites of ectopic CENP-A^{Cnp1} incorporation, but asymmetric AT-rich sequence features may play a secondary role. Furthermore, CENP-A^{Cnp1} may occupy additional sites within the central domains and the increased occupancy leads to changes in chromatin extraction that could reflect structural changes.

Chapter 5: A high-resolution sequence-based map of CENP-A/kinetochore interactions.

Conservation of strongly positioned CENP-A^{Cnp1} nucleosomes in fission yeasts indicates that positioning may be important for optimal centromere organisation. This could reflect the structure of interactions between CENP-A^{Cnp1} and the kinetochore or binding of a protein complex to the AT-rich DNA. The aim of this chapter was to investigate whether any proteins or complexes could be identified that make contacts with the asymmetric AT-rich DNA within the central domains and determine how kinetochore interactions were mediated at the highly positioned CENP-A^{Cnp1} nucleosomes. The analyses presented show that no kinetochore protein tested makes direct or indirect contacts with the AT-rich DNA, although the main candidate complex, the origin recognition complex, could not be tested at this time. However, this data supports a model where interactions with the kinetochore are mediated at all highly positioned CENP-A^{Cnp1} occupied sites, although within the population different sites would be occupied by CENP-A^{Cnp1} nucleosomes in different cells.

Chapter 2: Materials and methods

2.1 Growth media

PMG agar (500ml):

Pthallic acid	1.5 g
Na_2HPO_4	1.1 g
L-glutamic acid	1.875 g
D-glucose anhydrous	10 g
Salts 50x	10 ml
Minerals 10,0000x	0.05 ml
Vitamins 1000x	0.5 ml
Agar (OCOID)	10 g

PMG liquid (500ml):

Pthallic acid	1.5 g
Na_2HPO_4	1.1 g
L-glutamic acid	1.875 g
D-glucose anhydrous	10 g
Salts 50x	10 ml
Minerals 10,0000x	0.05 ml
Vitamins 1000x	0.5 ml

YES agar (500ml):

Yeast extract	2.5 g
D-glucose anhydrous	15 g
Adenine	0.1125 g
Histidine	0.1125 g
Leucine	0.1125 g
Uracil	0.1125 g
Lysine	0.1125 g
Agar (OXOID)	10 g

YES liquid (500ml):

Yeast extract	2.5 g

D-glucose anhydrous	15 g
Adenine	0.1125 g
Histidine	0.1125 g
Leucine	0.1125 g
Uracil	0.1125 g
Lysine	0.1125 g

Vitamins 1000x (100ml):

Pantothenic acid	0.5 g
Nicotinic acid	1 g
Inositol	1 g
Biotin	1 mg

Minerals 10,000x (100ml):

Boric acid	5 g
$MnSO_4$	4 g
$ZnSo_4$	4 g
$FeCl_2$ $6H_2O$	2 g
Molybdic acid	1.6 g
$CuSO_4$ $5H_2O$	0.4 g
Citric acid	10 g

Salts 50x:

Magnesium chloride	53.5 g
Calcium chloride	1 g
Potassium chloride	50 g
di-sodium sulphate	2 g

Supplements:

Adenine 50x (Sigma)	5 g/L
Arginine 100x (Sigma)	10 g/L
Histidine 100x (Sigma)	10 g/L
Uracil 20x (Sigma)	2 g/L
Leucine 100x (Sigma)	10 g/L

2.2 Chromatin-Immunoprecipitation (ChIP)

A 150 ml culture of cells was grown to a density of $1x10^7$ cells/ml in a shaking incubator at 32 °C. The density of cells was determined by measuring 100 µl of cells diluted in 10 ml of Isoton II using a Beckman Z2 Particle Count and Size Analyzer. Cells were transferred to 50ml falcon tubes and fixed in a fume hood with 1% formaldehyde for 15 minutes. Fixation was stopped with the addition of $1/20^{th}$ volume 2.5 M glycine. Cells were centrifuged for 2 minutes at 2500 rpm. Cells were washed twice with 20-40 ml PBS buffer. Cells were frozen at -80°C for further processing.

Cells were resuspended and lysed in PEMS with zymolase-100T at a density of 0.4 mg zymolase and $1x10^8$ cells per ml of PEMS. Cells were incubated for 1 hour at 37°C in a shaking incubator. Cells were centrifuged for 2 minutes.

Cells were resuspended in a 1.5ml eppendorf tube in 1ml lysis buffer with the addition of 1% protease inhibitor and 1% 2 mM PMSF by volume. Cells were sonicated using a BioRuptor for 20 minutes on high. Cells were centrifuged at 4°C for 20 minutes at 13000 rpm and the supernatant was transferred to a new tube. Protein G agarose beads or Dynabeads were washed with lysis buffer twice, centrifuged at 4°C for 1 minute at 5000rpm or placed on a magnet between washes, and resuspended in their original volume. If agarose beads were used then the crude lysates were pre-cleared by incubation with 40ul protein G agarose beads at 4°C for 10 minutes on a rotating wheel. 50ul of the crude lysate was taken for use as a control and processed directly for Proteinase K digestion with all subsequent steps being the same. The remaining sample was incubated with 50 ul anti-CENP-A^{Cnp1} serum or 7.5 ul anti-GFP antibody and 125 ul of beads overnight at 4°C on a rotating wheel.

Samples were washed with 1 ml lysis buffer, lysis buffer with 0.5 M NaCl (incubated on a rotating wheel for 10 minutes at 4°C), wash buffer (incubated on a rotating wheel for 10 minutes at 4°C), and TE at pH8. Between washes samples were either centrifuged at 2000 rpm at 4°C for 1 minute or placed on a magnet and resuspended by inverting. Samples were resuspended in 170 ul TES with 50 ul of 10 mg/ml proteinase K, vortexed, and incubated for 4 hours at 65°C.

Following incubation samples were vortexed and centrifuged or placed on a magnet. The supernatant was transferred to 1.3 ml buffer PB (Qiagen) in a new 2 ml eppendorf tube. For protein G agarose beads 100 ul TES was added to the sample, incubated for 5 minutes at 65°C, and transferred to the buffer PB to retrieve additional supernatant. Supernatant was transferred to a Qiagen PCR spin column 750 ul at a time and centrifuged to discard the supernatant. Samples were washed with 750 ul or 70% ethanol, centrifuged, and the flowthrough discarded. Further flowthrough was removed by knocking the collection tube on a paper towel. The column lids were cut off and the columns were centrifuged for 4 minutes at 13000 rpm. The spin basket was transferred to a new 1.5 ml eppendorf tube. 20 ul buffer EB was added to the spin basket and samples were incubated to 5 minutes. Samples were centrifuged for 30 seconds at 13000 rpm and frozen at -20°C.

Samples were subsequently analyzed by qPCR. Control samples were diluted 5 ul per 495 ul H_20 and ChIP samples were diluted 5ul/95ul H_20. A qPCR master mix was prepared with 5 ul KAPA SYBR FAST Master Mix Universal 2X qPCR Master Mix, 0.2 ul PCR forward primer, 0.2 ul PCR reverse primer, and 2.1 ul H_20 per sample. 3 ul of diluted sample and 7 ul of the qPCR master mix were aliquoted in each well in a 365 well PCR plate.

Lysis Buffer: 50 mM Hepes -KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Sodium Deoxycholate (DOC)
Lysis buffer w/ 0.5 M NaCl: Replace 140mM NaCl w/ 500mM
Wash Buffer:10 mM Tris-HCl, pH8, 0.25 M LiCl, 0.5% NP-40 (IGEPAL CA630) , 0.5% Sodium deoxycholate (DOC), 1 mM EDTA
TE: 10 mM Tris-HCl, pH8, 1 mM EDTA
TES: TE+1% SDS

2.3 Illumina library preparation

ChIP DNA was quantified by Qubit. 1-20 ng of input DNA was blunt ended by the addition of 1 mM dNPTs, 1 ul blunt enzyme mix, 5 ul 10X blunting buffer, and dH_20 in a total volume of 50 ul. Samples were incubated for 45 minutes at room temperature. DNA was purified by a 1.6:1 Ampure bead purification and eluted in 27.7 ul in dH_20.

DNA was A-tailed by the addition of 3.3 ul Buffer 2 (NEB), 1 ul 10 mM dATP, and 1 ul 5000 U/ml klenow (exo-). DNA was incubated on hot blocks at 37°C for 30 minutes, 75°C for 5 minutes to heat inactivate the enzyme, and 5 minutes on ice. DNA was immediately processed for adapter ligation. Adapters were ligated by the addition of 35 ul 2x Rapid DNA ligation buffer, 0.5 mM internally barcoded NEXTFLEX adaptors, and 1 ul quick T4 DNA ligase. DNA was incubated for 15 minutes at room temperature and purified via a 1:1 Ampure bead purification. DNA was eluted in 20 ul dH_20.

DNA was amplified by PCR to obtain libraries for sequencing by the addition of 10 ul sample DNA, 1 ul 10 mM dNTPs, 10 ul 5x Phusion HF buffer, 1 ul PCR primer mix, 1.5 ul DMSO, 0.5 ul Phusion polymerase, and dH_20 to a total volume of 50 ul. DNA was denatured at 98°C for 30 seconds and amplified by 16 cycles of denaturing at 98°C for 10 seconds, primer annealing at 65°C for 30 seconds, and extension at 72°C for 30 seconds, followed by a final extension at 72°C for 5 minutes. DNA was purified and size selected with a 0.7:1 Ampure bead purification with 1:1 Ampure bead cleanup and an additional 1:1 Ampure bead purification. Libraries were eluted in 20 ul. DNA concentration was measured by Qubit and library quality was assessed using a Bioanalyzer. Libraries were sent to Ark Genomics, Edinburgh for 100 bp paired end sequencing on a HiSeq2000.

Blunting Buffer: 100 mM Tris-HCl, 50 mM NaCl, 10 mM $MgCl_2$, 0.025% Triton X-100, 5 mM dithiothreitol, pH 7.5 at 25°C
Blunt Enymze mix: T4 DNA polymerase, T4 Polynucleotide Kinase, 100 mM KCl, 10 mM Tris-HCl (pH 7.4), 0.1 mM EDTA, 1 mM dithiothreitol, 0.1% Triton X-100 and 50% Glycerol
Quick Ligation Reaction Buffer: 66mM Tris-HCl, 10 mM $MgCl_2$, 1mM Dithiothreitol, 1 mM ATP, 7.5% Polyethylene glycol (PEG6000), pH 7.6 @ 25°C

2.4 General data analysis

Sequencing quality was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped to the *S. pombe* genome (*S. pombe* 972h$^-$ assembly ef2, 2007) using bowtie2 (Langmead and Salzberg, 2012; Langmead et al., 2009). Read pairs that mapped concordantly with a fragment size less than 250 bp were output to a file in SAM format (Li et al., 2009). Multiple maximum fragment sizes were tested but values ranging from 150 to 300 bp had minimal affect on the results since fragment sizes

were relatively tightly distributed around 150 bp. Mapped pairs were converted to BAM file format and sorted using SAMtools (Li et al., 2009). Peak calling and WIG format file generation was performed using MACS (Zhang et al., 2008b), keeping all duplicate fragments. The tight size selection and high enrichment over a small domain in some samples results in duplicate reads that are most likely not due to PCR artefacts. In agreement with this, FastQC finds no over-represented sequences and instead indicates that many read pairs are present more than once, which is consistent with very high enrichment of centromeres, and duplicate reads are observed primarily for samples with very high centromeric enrichment. Subpeaks were called using PeakSplitter with a valley depth cutoff of 0.6 (http://www.ebi.ac.uk/bertone/software/PeakSplitter_Cpp_usage.txt).

Subsequent general analyses were done using custom scripts I wrote in R (Team, 2013). Fragment coverage was calculated at per base resolution, normalised relative to millions of read pairs mapping, and the IP/IN ratio was calculated. Local AT-content was calculated genome-wide, at peaks, and at gaps at per-base resolution using a 147 bp sliding window using rollapply {zoo} (http://svitsrv25.epfl.ch/R-doc/library/zoo/html/rollapply.html). 2D density histograms were made using his2d {gplots} (http://svitsrv25.epfl.ch/R-doc/library/gplots/html/hist2d.html) and scatter plots were made using scatter.smooth {stats} (https://stat.ethz.ch/R-manual/R-patched/library/stats/html/scatter.smooth.html). Coverage bias due to AT-content was calculated at 10,000 random sites at the central domain and genome-wide for coverage of the IP, IN, and IP/IN ratio. Enrichment peaks were defined as +/-73 bp from the peak of enrichment and gaps were defined as regions outside these peak boundaries. Genome-wide *S. pombe* nucleosome occupancy prediction probabilities were downloaded from the Segal lab (http://genie.weizmann.ac.il/software/nucleo_genomes.html). Fragment coverage was plotted across the chromosomes and at regions of interest along with a track of annotated features downloaded from PomBase (Wood et al., 2012).

2.5 Analysis of CENP-A^{Cnp1} over-expression

Analyses were done using a custom script I wrote in R. Per base coverage scaled to the mapped reads per million was averaged between replicates for the IP and IN of CENP-A^{Cnp1} over-expression (OE) and CENP-A^{Cnp1} wild-type (WT) expression.

To measure changes in chromatin structure and extraction bias following CENP-A^{Cnp1} OE a test statistic

$$\boldsymbol{R_{IN}} = \frac{\boldsymbol{Coverage(OE_{IN})}}{\boldsymbol{Coverage(WT_{IN})}}$$

was calculated as the ratio between OE an WT IN coverage. To measure ectopic incorporation a test statistic

$$\boldsymbol{R_{IP}} = \frac{\boldsymbol{Coverage(OE_{IP})}}{\boldsymbol{Coverage(WT_{IP})}}$$

was calculated as the ratio between OE an WT IP coverage. The 1% highest and lowest R_{IN} values were used to set p-value significance thresholds for determining statistically significant higher and lower coverage following CENP-A^{Cnp1} OE, respectively. These p-value thresholds were applied to the R_{IP} values to determine regions of ectopic CENP-A^{Cnp1} incorporation.

The number of significant sites and the median R_{IP} value was calculated was calculated for 1 kb and 10 kb windows across the genome to search for regions of ectopic CENP-A^{Cnp1} incorporation. R_{IP} and R_{IN} values were plotted across chromosomes and at regions of interest along with a track of annotated features downloaded from PomBase (Wood et al., 2012). Violin plots were made using vioplot {vioplot} (http://cran.r-project.org/web/packages/vioplot/).

2.6 Analysis comparing kinetochore ChIP-Seq samples

Analyses were done using a custom script I wrote in R. Per base coverage scaled to the number of millions of reads mapping was determined for the IP, IN, and IP/IN ratio for all ChIP-Seq data sets. Pairwise correlations were done using the pairs {graphics} (http://stat.ethz.ch/R-manual/R-devel/library/graphics/html/pairs.html) function. Pairwise comparisons were calculated for 10,000 random sites within the central domains and genome-wide between all data sets using a non-parametric Kendall tau rank correlation coefficient. The number of peaks shared between data sets was calculated with a maximum distance between shared peaks ranging from 0 bp to 100 bp by intervals of 5 bp and visualised using venn {gplots] (http://svitsrv25.epfl.ch/R-doc/library/gplots/html/venn.html).

2.7 Nextflex barcodes:

Adapter	Index Sequence	Barcode Sequence
HT Adapter 1	AACGTGATATCTCGTATGCCGTCTTCTGCTTG	AACGTGAT
HT Adapter 2	AAACATCGATCTCGTATGCCGTCTTCTGCTTG	AAACATCG
HT Adapter 3	ATGCCTAAATCTCGTATGCCGTCTTCTGCTTG	ATGCCTAA
HT Adapter 4	AGTGGTCAATCTCGTATGCCGTCTTCTGCTTG	AGTGGTCA
HT Adapter 5	ACCACTGTATCTCGTATGCCGTCTTCTGCTTG	ACCACTGT
HT Adapter 6	ACATTGGCATCTCGTATGCCGTCTTCTGCTTG	ACATTGGC
HT Adapter 7	CAGATCTGATCTCGTATGCCGTCTTCTGCTTG	CAGATCTG
HT Adapter 8	CATCAAGTATCTCGTATGCCGTCTTCTGCTTG	CATCAAGT
HT Adapter 9	CGCTGATCATCTCGTATGCCGTCTTCTGCTTG	CGCTGATC
HT Adapter 10	ACAAGCTAATCTCGTATGCCGTCTTCTGCTTG	ACAAGCTA
HT Adapter 11	CTGTAGCCATCTCGTATGCCGTCTTCTGCTTG	CTGTAGCC
HT Adapter 12	AGTACAAGATCTCGTATGCCGTCTTCTGCTTG	AGTACAAG
HT Adapter 13	AACAACCAATCTCGTATGCCGTCTTCTGCTTG	AACAACCA
HT Adapter 14	AACCGAGAATCTCGTATGCCGTCTTCTGCTTG	AACCGAGA
HT Adapter 15	AACGCTTAATCTCGTATGCCGTCTTCTGCTTG	AACGCTTA
HT Adapter 16	AAGACGGAATCTCGTATGCCGTCTTCTGCTTG	AAGACGGA
HT Adapter 17	AAGGTACAATCTCGTATGCCGTCTTCTGCTTG	AAGGTACA
HT Adapter 18	ACACAGAAATCTCGTATGCCGTCTTCTGCTTG	ACACAGAA
HT Adapter 19	ACAGCAGAATCTCGTATGCCGTCTTCTGCTTG	ACAGCAGA
HT Adapter 20	ACCTCCAAATCTCGTATGCCGTCTTCTGCTTG	ACCTCCAA
HT Adapter 21	ACGCTCGAATCTCGTATGCCGTCTTCTGCTTG	ACGCTCGA
HT Adapter 22	ACGTATCAATCTCGTATGCCGTCTTCTGCTTG	ACGTATCA
HT Adapter 23	ACTATGCAATCTCGTATGCCGTCTTCTGCTTG	ACTATGCA
HT Adapter 24	AGAGTCAAATCTCGTATGCCGTCTTCTGCTTG	AGAGTCAA
HT Adapter 25	AGATCGCAATCTCGTATGCCGTCTTCTGCTTG	AGATCGCA
HT Adapter 26	AGCAGGAAATCTCGTATGCCGTCTTCTGCTTG	AGCAGGAA
HT Adapter 27	AGTCACTAATCTCGTATGCCGTCTTCTGCTTG	AGTCACTA
HT Adapter 28	ATCCTGTAATCTCGTATGCCGTCTTCTGCTTG	ATCCTGTA
HT Adapter 29	ATTGAGGAATCTCGTATGCCGTCTTCTGCTTG	ATTGAGGA
HT Adapter 30	CAACCACAATCTCGTATGCCGTCTTCTGCTTG	CAACCACA
HT Adapter 31	CAAGACTAATCTCGTATGCCGTCTTCTGCTTG	GACTAGTA
HT Adapter 32	CAATGGAAATCTCGTATGCCGTCTTCTGCTTG	CAATGGAA

HT Adapter 33	CACTTCGAATCTCGTATGCCGTCTTCTGCTTG	CACTTCGA
HT Adapter 34	CAGCGTTAATCTCGTATGCCGTCTTCTGCTTG	CAGCGTTA
HT Adapter 35	CATACCAAATCTCGTATGCCGTCTTCTGCTTG	CATACCAA
HT Adapter 36	CCAGTTCAATCTCGTATGCCGTCTTCTGCTTG	CCAGTTCA
HT Adapter 37	CCGAAGTAATCTCGTATGCCGTCTTCTGCTTG	CCGAAGTA
HT Adapter 38	CCGTGAGAATCTCGTATGCCGTCTTCTGCTTG	CCGTGAGA
HT Adapter 39	CCTCCTGAATCTCGTATGCCGTCTTCTGCTTG	CCTCCTGA
HT Adapter 40	CGAACTTAATCTCGTATGCCGTCTTCTGCTTG	CGAACTTA
HT Adapter 41	CGACTGGAATCTCGTATGCCGTCTTCTGCTTG	CGACTGGA
HT Adapter 42	CGCATACAATCTCGTATGCCGTCTTCTGCTTG	CGCATACA
HT Adapter 43	CTCAATGAATCTCGTATGCCGTCTTCTGCTTG	CTCAATGA
HT Adapter 44	CTGAGCCAATCTCGTATGCCGTCTTCTGCTTG	CTGAGCCA
HT Adapter 45	CTGGCATAATCTCGTATGCCGTCTTCTGCTTG	CTGGCATA
HT Adapter 46	GAATCTGAATCTCGTATGCCGTCTTCTGCTTG	GAATCTGA
HT Adapter 47	GACTAGTAATCTCGTATGCCGTCTTCTGCTTG	CAAGACTA
HT Adapter 48	GAGCTGAAATCTCGTATGCCGTCTTCTGCTTG	GAGCTGAA
HT Adapter 49	GATAGACAATCTCGTATGCCGTCTTCTGCTTG	GATAGACA
HT Adapter 50	GCCACATAATCTCGTATGCCGTCTTCTGCTTG	GCCACATA
HT Adapter 51	GCGAGTAAATCTCGTATGCCGTCTTCTGCTTG	GCGAGTAA
HT Adapter 52	GCTAACGAATCTCGTATGCCGTCTTCTGCTTG	GCTAACGA
HT Adapter 53	GCTCGGTAATCTCGTATGCCGTCTTCTGCTTG	GCTCGGTA
HT Adapter 54	GGAGAACAATCTCGTATGCCGTCTTCTGCTTG	GGAGAACA
HT Adapter 55	GGTGCGAAATCTCGTATGCCGTCTTCTGCTTG	GGTGCGAA
HT Adapter 56	GTACGCAAATCTCGTATGCCGTCTTCTGCTTG	GTACGCAA
HT Adapter 57	GTCGTAGAATCTCGTATGCCGTCTTCTGCTTG	GTCGTAGA
HT Adapter 58	GTCTGTCAATCTCGTATGCCGTCTTCTGCTTG	GTCTGTCA
HT Adapter 59	GTGTTCTAATCTCGTATGCCGTCTTCTGCTTG	GTGTTCTA
HT Adapter 60	TAGGATGAATCTCGTATGCCGTCTTCTGCTTG	TAGGATGA
HT Adapter 61	TATCAGCAATCTCGTATGCCGTCTTCTGCTTG	TATCAGCA
HT Adapter 62	TCCGTCTAATCTCGTATGCCGTCTTCTGCTTG	TCCGTCTA
HT Adapter 63	TCTTCACAATCTCGTATGCCGTCTTCTGCTTG	TCTTCACA
HT Adapter 64	TGAAGAGAATCTCGTATGCCGTCTTCTGCTTG	TGAAGAGA
HT Adapter 65	TGGAACAAATCTCGTATGCCGTCTTCTGCTTG	TGGAACAA
HT Adapter 66	TGGCTTCAATCTCGTATGCCGTCTTCTGCTTG	TGGCTTCA

HT Adapter 67	TGGTGGTAATCTCGTATGCCGTCTTCTGCTTG	TGGTGGTA
HT Adapter 68	TTCACGCAATCTCGTATGCCGTCTTCTGCTTG	TTCACGCA
HT Adapter 69	AACTCACCATCTCGTATGCCGTCTTCTGCTTG	AACTCACC
HT Adapter 70	AAGAGATCATCTCGTATGCCGTCTTCTGCTTG	AAGAGATC
HT Adapter 71	AAGGACACATCTCGTATGCCGTCTTCTGCTTG	AAGGACAC
HT Adapter 72	AATCCGTCATCTCGTATGCCGTCTTCTGCTTG	AATCCGTC
HT Adapter 73	AATGTTGCATCTCGTATGCCGTCTTCTGCTTG	AATGTTGC
HT Adapter 74	ACACGACCATCTCGTATGCCGTCTTCTGCTTG	ACACGACC
HT Adapter 75	ACAGATTCATCTCGTATGCCGTCTTCTGCTTG	ACAGATTC
HT Adapter 76	AGATGTACATCTCGTATGCCGTCTTCTGCTTG	AGATGTAC
HT Adapter 77	AGCACCTCATCTCGTATGCCGTCTTCTGCTTG	AGCACCTC
HT Adapter 78	AGCCATGCATCTCGTATGCCGTCTTCTGCTTG	AGCCATGC
HT Adapter 79	AGGCTAACATCTCGTATGCCGTCTTCTGCTTG	AGGCTAAC
HT Adapter 80	ATAGCGACATCTCGTATGCCGTCTTCTGCTTG	ATAGCGAC
HT Adapter 81	ATCATTCCATCTCGTATGCCGTCTTCTGCTTG	ATCATTCC
HT Adapter 82	ATTGGCTCATCTCGTATGCCGTCTTCTGCTTG	ATTGGCTC
HT Adapter 83	CAAGGAGCATCTCGTATGCCGTCTTCTGCTTG	CAAGGAGC
HT Adapter 84	CACCTTACATCTCGTATGCCGTCTTCTGCTTG	CACCTTAC
HT Adapter 85	CCATCCTCATCTCGTATGCCGTCTTCTGCTTG	CCATCCTC
HT Adapter 86	CCGACAACATCTCGTATGCCGTCTTCTGCTTG	CCGACAAC
HT Adapter 87	CCTAATCCATCTCGTATGCCGTCTTCTGCTTG	CCTAATCC
HT Adapter 88	CCTCTATCATCTCGTATGCCGTCTTCTGCTTG	CCTCTATC
HT Adapter 89	CGACACACATCTCGTATGCCGTCTTCTGCTTG	CGACACAC
HT Adapter 90	CGGATTGCATCTCGTATGCCGTCTTCTGCTTG	CGGATTGC
HT Adapter 91	CTAAGGTCATCTCGTATGCCGTCTTCTGCTTG	CTAAGGTC
HT Adapter 92	GAACAGGCATCTCGTATGCCGTCTTCTGCTTG	GAACAGGC
HT Adapter 93	GACAGTGCATCTCGTATGCCGTCTTCTGCTTG	GACAGTGC
HT Adapter 94	GAGTTAGCATCTCGTATGCCGTCTTCTGCTTG	GAGTTAGC
HT Adapter 95	GATGAATCATCTCGTATGCCGTCTTCTGCTTG	GATGAATC
HT Adapter 96	GCCAAGACATCTCGTATGCCGTCTTCTGCTTG	GCCAAGAC

2.8 Strains used

Strain	Genotype
143	*S. pombe* h^-

3826	*S. pombe h⁻ LEU 1 mis12+GFP-LEU2+*
4569	*S. pombe h⁻ Ndc80-GFP-Kan^R bub1-HA-ura4+ leu1-32 ura4D18?*
5147	*S. pombe h⁻ cnp3GFP::Kan^{R}ade6-210 ura4D18 leu1-32 his3D1 arg3D4*
5929	*S. pombe h⁺ sim1-GFP*
A86	*S. pombe h⁻ leu1 ura4 mis14-GFP[ura4+]*
A90	*S. pombe h⁻ leu1 ura4 mis16-GFP[ura4+]*
A94	*S. pombe h⁻ leu1 mis18-GFP[LEU2+]*
A244	*S. pombe h? cnp20-GFP-Kan^R Rint:ura4 leu1-32 ura4DS/E his3D ade6-210 arg3D4*
A325	*S. pombe h⁻ ars1(MluI)::pREP3X-leu2Sc leu1-32 (#100)*
A334	*S. pombe h⁺ ars1(MluI)::pREP41XCnp1-leu2Sc leu1-32 (#109)*
A6372	*S. pombe h⁻ leu1 ura4 ∆cen1::pADH1-loxP-Kan^R cd39 (tel1L neocentromere)*
A6969	*S. octosporus h90*
A6972	*S. cryophillus h90*
A7984	*S. pombe h⁺ leu1-32 ura4-D18 ade6-M210 his3-D1 arg3-D4 eic2:GFP-Kan^R*
A7988	*S. pombe h⁺ leu1-32 ura4-D18 ade6-M210 his3-D1 arg3-D4 eic1:GFP-Kan^R*
A9589	*h⁻ GFP-cnp1 NAT*

2.9 ChIP antibodies used

GFP Polyclonal antibody, rabbit IgG fraction, A11122 (Life Technologies)

anti-CENP-A^{Cnp1} serum sheep

Monoclonal H3K9me2 antibody, m5.1.1 (Gift from Takeshi Urano)

2.10 Primers used

ID	Info	Primer sequence
NT15	dg 108, F qPCR	TCCAAATGTCGCATGAACACTC
NT16	dg 108, R qPCR	
	Solexa Library	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATT
NT17	PCR primer 1.1	CCTGCTGAACCGCTCTTCCGATC*T
NT18	Solexa Library	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC

	PCR primer 2.1	TACACGACGCTCTTCCGATC*T
NT19	Solexa Library Adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATC*T
NT20	Solexa Library Adapter	[Phos]GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
HB 424	cc2 C, F qPCR	AAGCGCTAACTCGTTTAAGTGAA
HB 425	cc2 C, R qPCR	ACATGGCGTGGAAAGTCATC
HB 475	act1, F qPCR	CCG GCG AGA TCA AGA CGC AT
HB 476	act1, R qPCR	TAT GTT GCT ATT CAA GCT G
NFF	NEXTFLEX primer F	AATGATACGGCGACCACCGAGATCTACAC
NFR	NEXTFLEX primer R	CAAGCAGAAGACGGCATACGAGAT

2.11 ChIP-Seq data

File	Species	Sequencer	Experiment	# of reads	Additional info
NT01_J	*S. pombe*	Genepool	4569 IP		Ndc80-GFP chip-seq IP
NT02_1	*S. pombe*	Ark Genomics	5929 IP	2530658	Scm3-GFP Chip-seq IP
NT02_7	*S. pombe*	Ark Genomics	5929 IP	3617448	Scm3-GFP Chip-seq IN
NT04_1	*S. pombe*	Ark Genomics	3826 IP	2323507	Mis12-GFP ChIP-Seq IP
NT04_2	*S. pombe*	Ark Genomics	4569 IP	1977369	Ndc80-GFP chip-seq IP
NT04_3	*S. pombe*	Ark Genomics	5147 IP	3430878	Cnp3-GFP ChIP-Seq IP
NT04_4	*S. pombe*	Ark Genomics	5929 IP	3008540	Scm3-GFP Chip-seq IP

NT04_5	*S. pombe*	Ark Genomics	A86 IP	2753575	Mis14-GFP ChIP-Seq IP
NT04_6	*S. pombe*	Ark Genomics	A90 IP	3692731	Mis16-GFP ChIP-Seq IP
NT04_7	*S. pombe*	Ark Genomics	A94 IP	2060548	Mis18-GFP ChIP-Seq IP
NT04_8	*S. pombe*	Ark Genomics	A244 IP	2909593	Cnp20-GFP ChIP-Seq IP
NT04_9	*S. pombe*	Ark Genomics	A7984 IP	3930432	Eic2-GFP ChIP-Seq IP
NT04_10	*S. pombe*	Ark Genomics	A7988 IP	3120510	Eic1-GFP ChIP-Seq IP
NT04_11	*S. pombe*	Ark Genomics	3826 IN	4497950	Mis12-GFP ChIP-Seq IN
NT04_12	*S. pombe*	Ark Genomics	4569 IN	4956220	Ndc80-GFP chip-seq IN
NT04_13	*S. pombe*	Ark Genomics	5147 IN	5005268	Cnp3-GFP ChIP-Seq IN
NT04_14	*S. pombe*	Ark Genomics	5929 IN	5722657	Scm3-GFP Chip-seq IN
NT04_15	*S. pombe*	Ark Genomics	A86 IN	5539295	Mis14-GFP ChIP-Seq IN
NT04_16	*S. pombe*	Ark Genomics	A90 IN	4717863	Mis16-GFP ChIP-Seq IN
NT04_17	*S. pombe*	Ark Genomics	A94 IN	6363904	Mis18-GFP ChIP-Seq IN
NT04_18	*S. pombe*	Ark Genomics	A244 IN	4803427	Cnp20-GFP ChIP-Seq IN
NT04_19	*S. pombe*	Ark Genomics	A7984 IN	5306861	Eic2-GFP ChIP-Seq IN
NT04_20	*S. pombe*	Ark Genomics	A7988 IN	5390005	Eic1-GFP ChIP-Seq IN
NT06_1	*S. pombe*	Ark Genomics	143 HA IP	5246888	WT anti-HA ChIP-Seq IP

NT06_2	*S. pombe*	Ark Genomics	3826 IP	3777669	Mis12-GFP ChIP-Seq IP
NT06_4	*S. pombe*	Ark Genomics	5147 IP	5831976	Cnp3-GFP ChIP-Seq IP
NT06_7	*S. pombe*	Ark Genomics	5929 IP	7929579	Scm3-GFP ChIP-Seq IP
NT06_12	*S. pombe*	Ark Genomics	143 HA IN	6672482	WT anti-HA ChIP-Seq IN
NT06_13	*S. pombe*	Ark Genomics	3826 IN	5432137	Mis12-GFP ChIP-Seq IN
NT06_15	*S. pombe*	Ark Genomics	5147 IN	5251689	Cnp3-GFP ChIP-Seq IN
NT06_18	*S. pombe*	Ark Genomics	5929 IN	5981125	Scm3-GFP ChIP-Seq IN
N001_1	*S. pombe*	Ark Genomics	143 Cnp1 IP	1846906	WT Cnp1 IP
N001_2	*S. pombe*	Ark Genomics	143 GFP IP	2299922	WT mock IP
N001_4	*S. pombe*	Ark Genomics	A325 A Cnp1 IP	4612335	Cnp1 WT IP
N001_5	*S. pombe*	Ark Genomics	A325 A Cnp1 IP	4707432	Cnp1 WT IP
N001_6	*S. pombe*	Ark Genomics	A325 B Cnp1 IP	5723511	Cnp1 WT IP
N001_7	*S. pombe*	Ark Genomics	A325 B Cnp1 IP	3598962	Cnp1 WT IP
N001_8	*S. pombe*	Ark Genomics	A334 A Cnp1 IP	4024882	Cnp1 OE IP
N001_9	*S. pombe*	Ark Genomics	A334 A Cnp1 IP	7055320	Cnp1 OE IP
N001_10	*S. pombe*	Ark Genomics	A334 B Cnp1 IP	6056199	Cnp1 OE IP
N001_11	*S. pombe*	Ark Genomics	143 Cnp1 IN	5803546	WT Cnp1 IN

N001_12	*S. pombe*	Ark Genomics	143 GFP IN	6672482	WT mock IN
N001_14	*S. pombe*	Ark Genomics	A325 A Cnp1 IN	2422495	Cnp1 WT IN
N001_15	*S. pombe*	Ark Genomics	A325 A Cnp1 IN	5251689	Cnp1 WT IN
N001_16	*S. pombe*	Ark Genomics	A325 B Cnp1 IN	5290804	Cnp1 WT IN
N001_17	*S. pombe*	Ark Genomics	A325 B Cnp1 IN	3508456	Cnp1 OE IN
N001_18	*S. pombe*	Ark Genomics	A334 A Cnp1 IN	5981125	Cnp1 OE IN
N001_19	*S. pombe*	Ark Genomics	A334 A Cnp1 IN	6528925	Cnp1 OE IN
N001_20	*S. pombe*	Ark Genomics	A334 B Cnp1 IN	6237334	Cnp1 OE IN
N002_1	*S. pombe*	Ark Genomics	A6372 Cnp1 IP	1469903	cd39 Cnp1 IP
N002_2	*S. pombe*	Ark Genomics	A6372 H3K9me IP	5844070	cd39 H3K9 IP
N002_3	*S. pombe*	Ark Genomics	A6372 Cnp1 IP	3105345	cd39 Cnp1 IP
N002_4	*S. pombe*	Ark Genomics	A6372 H3K9me IP	5459465	cd39 H3K9 IP
N002_5	*S. octosporus*	Ark Genomics	A6969 Cnp1 IP	3275997	Cnp1 IP
N002_6	*S. octosporus*	Ark Genomics	A6969 H3K9me IP	4012175	H3K9 IP
N002_7	*S. cryophilus*	Ark Genomics	A6972 H3K9me IP	7295375	H3K9 IP
N002_8	*S. cryophilus*	Ark Genomics	A6972 Cnp1 IP	3855782	Cnp1 IP
N002_9	*S. pombe*	Ark Genomics	A334 B Cnp1 IP	7772166	Cnp1 OE IP

N002_10	*S. pombe*	Ark Genomics	A6372 Cnp1 IN	3705419	cd39 Cnp1 IN
N002_11	*S. pombe*	Ark Genomics	A6372 H3K9me IN	6528255	cd39 K9 IN
N002_12	*S. pombe*	Ark Genomics	A6372 Cnp1 IN	4321576	cd39 Cnp1 IN
N002_13	*S. pombe*	Ark Genomics	A6372 H3K9me IN	4981959	cd39 K9 IN
N002_14	*S. octosporus*	Ark Genomics	A6969 Cnp1 IN	6860021	Cnp1 IN
N002_15	*S. octosporus*	Ark Genomics	A6969 H3K9me IN	3423126	H3K9 IN
N002_16	*S. cryophilus*	Ark Genomics	A6972 H3K9me IN	5488448	H3K9 IN
N002_17	*S. cryophilus*	Ark Genomics	A6972 Cnp1 IN	6368701	Cnp1 IN
N002_18	*S. pombe*	Ark Genomics	A334 B Cnp1 IN	10538219	Cnp1 OE IN
N2290_1_1	*S. pombe*	Ark Genomics	143 H3 IN	6226272	WT H3 IN
N2290_1_4	*S. pombe*	Ark Genomics	A9581 IN	10141980	Mock IN agarose
N2290_1_5	*S. pombe*	Ark Genomics	A9581 IN	7821695	Mock IN dynabeads
N2290_1_6	*S. pombe*	Ark Genomics	A9587 IN	12148659	Ndc80-GFP ChIP-seq IN
N2290_1_7	*S. pombe*	Ark Genomics	A9588 IN	7999281	Cnp20-GFP ChIP-Seq IN
N2290_1_8	*S. pombe*	Ark Genomics	A9589 IN	9989253	GFP-Cnp1 ChIP-Seq IN
N2290_1_9	*S. pombe*	Ark Genomics	143 H3 IP	7221480	H3 IP
N2290_1_12	*S. pombe*	Ark Genomics	A9581 IP	6605032	Mock IP agarose IP

N2290_1_13	*S. pombe*	Ark Genomics	A9587 IP	8377885	Ndc80-GFP ChIP-seq IP
N2290_1_14	*S. pombe*	Ark Genomics	A9588 IP	8228995	Cnp20-GFP ChIP-Seq IP
N2290_1_15	*S. pombe*	Ark Genomics	A9589 IP	7052673	GFP-Cnp1 ChIP-Seq IP
N2290_2_4	*S. pombe*	Ark Genomics	A9581 IN	7643570	Mock agarose beads IN
N2290_2_5	*S. pombe*	Ark Genomics	A9581 IN	7499098	Mock Dynabeads IN
N2290_2_6	*S. pombe*	Ark Genomics	A9587 IN	10039662	Ndc80-GFP ChIP-seq IN
N2290_2_7	*S. pombe*	Ark Genomics	A9588 IN	5434914	Cnp20-GFP ChIP-Seq IN
N2290_2_8	*S. pombe*	Ark Genomics	A9589 IN	11753660	GFP-Cnp1 ChIP-Seq IN
N2290_2_12	*S. pombe*	Ark Genomics	A9581 IP	6280832	Mock IP agarose beads IP
N2290_2_13	*S. pombe*	Ark Genomics	A9581 IP	4807301	Mock IP Dynabeads IP
N2290_2_14	*S. pombe*	Ark Genomics	A9587 IP	7536208	Ndc80-GFP ChIP-seq IP
N2290_2_15	*S. pombe*	Ark Genomics	A9588 IP	9339259	Cnp20-GFP ChIP-Seq IP
N2290_2_16	*S. pombe*	Ark Genomics	A9589 IP	9676541	GFP-Cnp1 ChIP-Seq IP

Chapter 3: Conservation of CENP-A^{Cnp1} positioning sequence features in fission yeasts

3.1: Introduction

Despite the conservation of CENP-A and many kinetochore components between eukaryotic species, centromere sequences are rapidly evolving and often show no homology, even between closely related species (Henikoff et al., 2001). Centromere structure varies greatly between species, from a sequence dependent point centromere in budding yeast to regional centromeres tens and hundreds of kilobases in length in fission yeast and mammals, respectively (Fukagawa and Earnshaw, 2014). The diversity of centromere structures and sequences raises the question of how centromere identity is defined.

The centromere specific protein CENP-A is generally accepted as the epigenetic mark for centromere identity and centromeric sequences are not required for centromere activity. Nevertheless, certain sequences are correlated with centromere activity. Most directly, the point centromeres of the budding yeast *S. cerevisiae* share an ~125 bp arrangement of sequence motifs (Fitzgerald-Hayes et al., 1982). Even epigenetically determined centromeres are often associated with particular sequences, including centromeres that form on repeats in humans (Mitchell et al., 1985; Waye and Willard, 1989), chickens (Shang et al., 2010), and *Drosophila* (Sun et al., 2003). The fission yeast centromere contains a non-repetitive central core, but is also associated with flanking outer-repeat elements (Clarke and Baum, 1990; Clarke et al., 1986; Fishel et al., 1988; Hahnenberger et al., 1991; Murakami et al., 1991; Nakaseko et al., 1987). None of these sequences (besides the budding yeast centromere) are absolutely necessary centromere function, but they suggest that some aspect of these sequences contributes to optimal centromere activity.

If CENP-A nucleosomes have different sequence preferences than canonical H3 nucleosomes then centromere associated sequences may allow CENP-A to be targeted to and incorporated into centromere sequences with greater affinity. This possibility was investigated in a budding yeast *in vitro* nucleosome deposition assay (Visnapuu and Greene, 2009). Nucleosomes containing H2A, H2B, H4, and CENP-A^{Cse4} had similar sequence preferences as canonical nucleosomes, including a

preference against AT-rich sequences. The properties of a Scm3/H4/CENP-A^{Cse4} hexamer were also tested since the CENP-A chaperone Scm3 had been proposed to be incorporated into hexameric CENP-A^{Cse4} nucleosomes in budding yeast (Camahort et al., 2007; Mizuguchi et al., 2007; Stoler et al., 2007; Williams et al., 2009). An Scm3/H4/CENP-A^{Cse4} nucleosome showed no preference against more AT-rich sequences (Visnapuu and Greene, 2009). A second study similarly found that a reconstituted octameric CENP-A^{Cse4} nucleosome could not assemble efficiently on centromeric DNA, while a complex containing Scm3 could (Xiao et al., 2011). It was proposed that Scm3 is responsible for the targeting and deposition of CENP-A^{Cse4} nucleosomes at the budding yeast centromere (Visnapuu and Greene, 2009; Xiao et al., 2011). However, it was observed that CENP-A^{Cse4} is stably maintained at centromeres while Scm3 association is highly dynamic, essentially ruling out the possibility of a hexameric structure (Wisniewski et al., 2014).

Sequence features could also influence CENP-A positioning within the centromere. Analysis using tiling arrays of *S. pombe* the chromosome II centromere suggested that MNase digested CENP-A^{Cnp1} occupied discrete nucleosome peaks positioned with a preference against AT-rich sequences, though the centromere is more AT-rich than the genome-wide average (Song et al., 2008). These data were limited by the low spatial resolution and dynamic range inherent in tiling arrays. Additionally, CENP-A is known to be more sensitive to exonuclease activity than canonical nucleosomes (Conde e Silva et al., 2007) and MNase has sequence specific cleavage biases (Chung et al., 2010; Dingwall et al., 1981; Hörz and Altenburger, 1981). Therefore, occupancy of CENP-A^{Cnp1} positioning at fission yeast centromeres warrants further investigation. Research by H. Berger and M. Miell (unpublished, Allshire lab; Miell, Thesis 2013) sought to determine whether CENP-A^{Cnp1} nucleosomes differed from canonical H3 nucleosomes in their preference for sequence features in fission yeast and whether this could influence CENP-A^{Cnp1} occupancy within the centromere and specific targeting to the centromere. They found that CENP- A^{Cnp1} occupancy is determined, at least partially, in a sequence dependent manner with similar sequence preferences as canonical nucleosomes. Such analyses suggest that sequence preferences are not responsible for targeting CENP-A^{Cnp1} to centromeres to determine centromere identity, but are at least partially responsible for the occupancy patterns of CENP-A ^{Cnp1} within the central domain region of fission yeast centromeres.

These previous analyses focused on correlations that give insight into sequence preferences of H3 and CENP-A^{Cnp1} nucleosomes, but the analyses did not elucidate the specific sequence features affecting nucleosome occupancy at and within centromere DNA or the organisation of these features within the central domain. The current study aimed to identify the sequence features influencing CENP-A^{Cnp1} nucleosome occupancy within the centromere at high spatial resolution. Additionally, it was investigated how these sequence features are organised within the central domain and whether this organisation is conserved at the centromeres of related fission yeast species. CENP-A^{Cnp1} occupancy at neocentromeres was also analysed to determine which of these sequence features are strictly required for centromere activity at an evolutionarily young centromere.

3.2: Results

3.2.1 CENP-A^{Cnp1} is highly positioned in a sequence dependent manner

Published CENP-A^{Cnp1} ChIP-seq data indicated that CENP-A^{Cnp1} nucleosomes are highly positioned at discrete sites within the central domains (central core plus inverted imr repeats) of fission yeast centromeres, with large gaps between nucleosomes (Lando et al., 2012). Unpublished *in vitro* data suggests that this positioning is sequence dependent (H. Berger and M. Miell, Allshire lab). However, it remains unknown what sequence features mediate nucleosome positioning at these specific sites within the central domain and how these features are organised within centromeres.

I have reanalysed the previously published CENP-A^{Cnp1} ChIP-seq data from Lando et al. (2012) to determine how sequence features contribute to the spatial pattern of CENP-A ^{Cnp1} positioning at discrete sites within the centromere. This data was generated from either sonicated or MNase digested chromatin that was immunoprecipitated (IP) with anti-CENP-A^{Cnp1} antiserum and sent for high throughput Illumina sequencing, along with purified sheared or MNase digested input chromatin (IN).

The sonicated and MNase-digested CENP-A^{Cnp1} ChIP-seq data indicate that CENP-A^{Cnp1} is highly enriched at centromeres (Figure 3.1). The centromere of chromosome II is shown as a representative example when plotting coverage, but

all analyses in this thesis were done using all three centromeres and results are consistent between different centromeres. The sonicated ChIP-seq data is presented as the coverage of the IP relative to the IN. This controls for technical biases in the recovery of DNA and sequencing bias. The MNase data could not be normalised in this way because the MNase IN control had extremely low read coverage within centromere regions. Within the central domains the sonicated IP/IN ratios and MNase IP coverage are moderately correlated within the three central domains (τ=0.34 [I], 0.48 [II], 0.43 [III]) and both predict the presence of highly positioned CENP-A^{Cnp1} nucleosomes with large gaps between them. Discrete peaks of CENP-A^{Cnp1} occupancy were determined using MACS and PeakSplitter. Gaps between nucleosomes were defined as regions that were greater than 75 bp from the midpoint any peak. There is a median gap size of 508 bp (sonicated) and 392 bp (MNase) between CENP-A^{Cnp1} peaks (Figure 3.2). The difference in the median gap size is the result of a trade-off between peak calling power and spatial resolution. The MNase sample has higher spatial resolution so that adjacent peaks can be called with greater accuracy, but the highly variable coverage results in low coverage at some peaks. As such, neither analysis can be considered strictly accurate, but possible solutions to this problem are discussed later.

DNA sequence AT-content has been shown to have a significant influence on preferences for canonical H3 nucleosome occupancy. Generally, highly AT-rich sequences tend to be relatively nucleosome depleted with low levels of nucleosome occupancy both *in vivo* and *in vitro* (Hughes and Rando, 2014; Struhl and Segal, 2013). Therefore, AT-content was initially investigated as a sequence feature that could affect CENP-A^{Cnp1} occupancy. Local AT-content was calculated at base pair resolution across the genome using a 147 base pair sliding window. This window size was chosen to reflect the approximate amount of DNA wrapped by a canonical nucleosome and which would affect the probability of a nucleosome occupying a

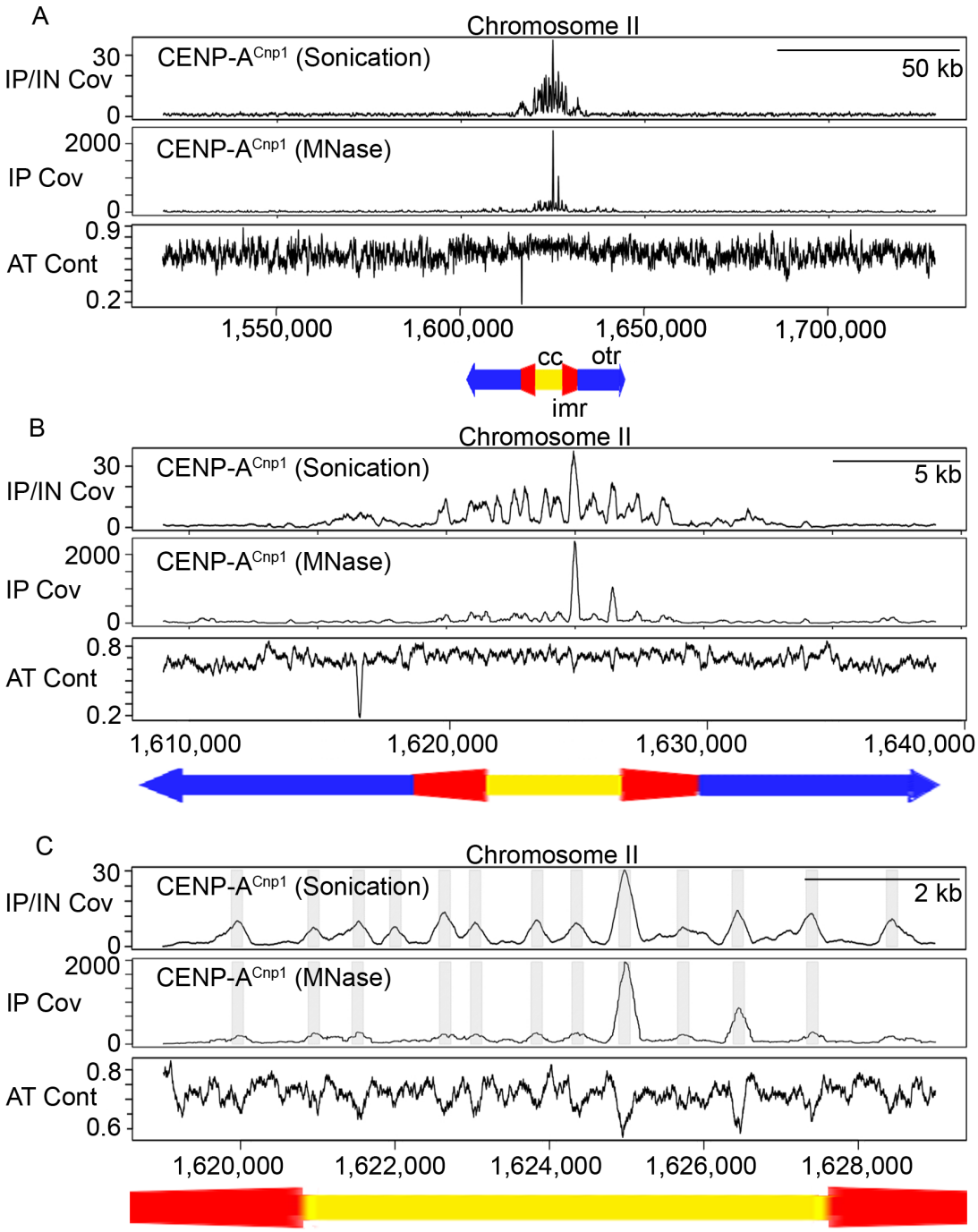


Figure 3.1. CENP-A^{Cnp1} is enriched at the centromere central domains. Sonicated and MNase CENP-A^{Cnp1} ChIP-seq data (Lando et al. 2012) mapped to the fission yeast genome and visualised at central domain of the chromosome II centromere, including (A) 100 kb, (B) 10 kb, and (C) 0 kb of flanking sequences. Both sonicated and MNase digested CENP-A^{Cnp1} ChIP-seq data show that CENP-A^{Cnp1} is highly enriched specifically at the central domain, as previously observed. CENP-A^{Cnp1} is enriched at discrete peaks (indicated by grey boxes) within the central domain with large gaps between peaks.

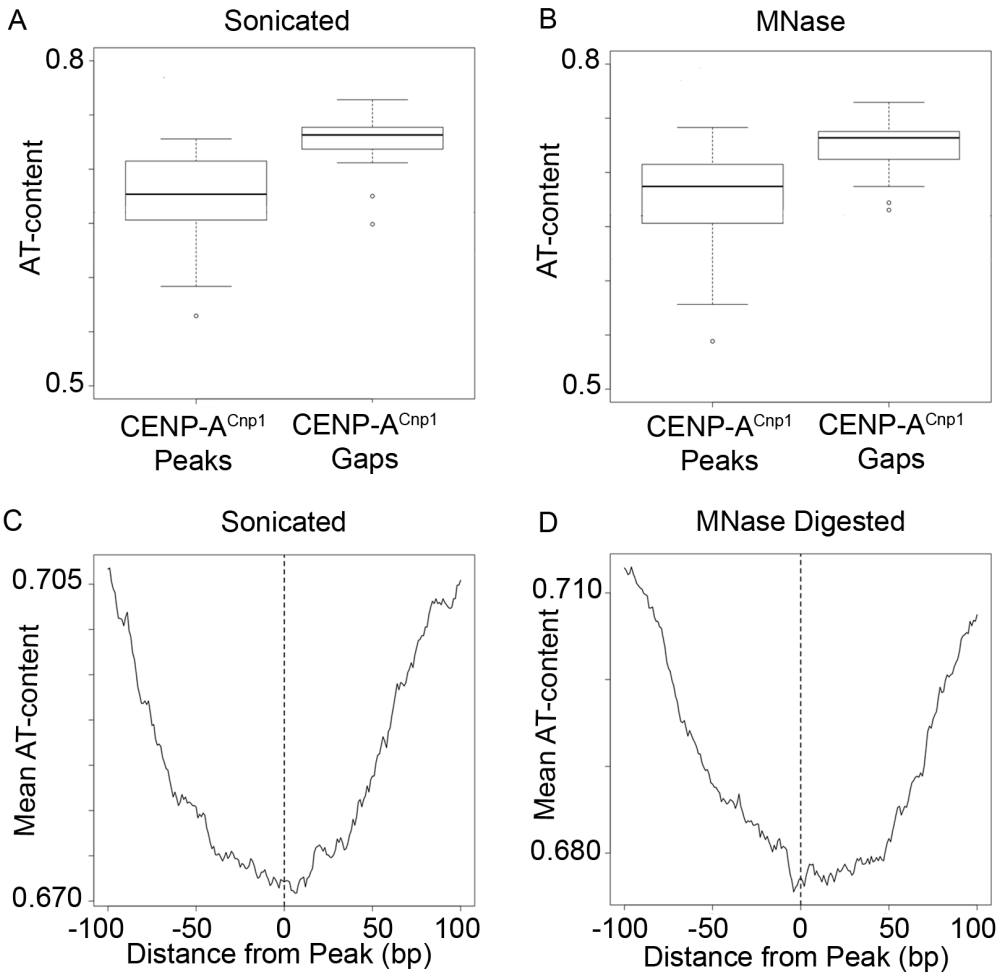


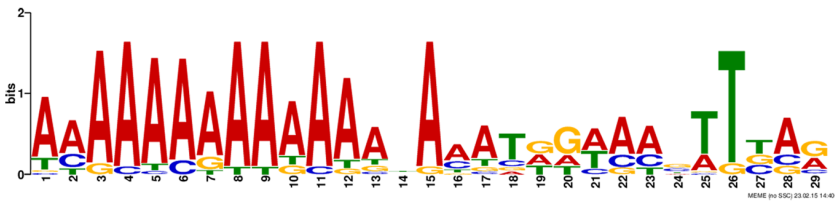
Figure 3.2. CENP-A^{Cnp1} is highly positioned by sequence features. (A) CENP-A^{Cnp1} occupied sites are separated by large inter-nucleosomal gaps (median= 508 bp [sonicated], 392 bp [MNase]). (B) The sites occupied by CENP-A^{Cnp1} are less AT-rich (mean=67.0% [sonicated], 67.7% [MNase]) than the gaps between CENP-A^{Cnp1} (mean=72.9% [sonicated],72.7% [MNase]). This is in comparison to a genome-wide AT-content of 63.9%. (C) CENP-A^{Cnp1} is positioned at local minima of AT-content with increasing AT-content as the distance from the centre of the peak increases.

given site. Human CENP-A containing nucleosomes have a smaller MNase resistant footprint and wrap less DNA directly with terminal DNA wrapped only partially (Tachiwana et al., 2011b). This has also been found to be the case for recombinant *S. pombe* nucleosome (Miell, Thesis 2013). However, these less stable terminal interactions may still influence nucleosome occupancy and positioning. Additionally, it is of particular interest to determine sequence features that influence occupancy of H3 nucleosomes and other DNA interacting complexes within the central domains. Therefore 147 bp was used as the window size for these analyses. The window size is only used for the analyses looking at the relationship between AT-content and distance from the peak position and altering the window size has minimal effect on the results.

On average *S. pombe* centromere DNA is 4.3 to 6.2% more AT-rich (68.8% [I], 68.4% [II], 70.0% [III]) than chromosomal DNA (63.9% [I], 64.1% [II], 63.8% [III]). The central domains of centromeres, the regions of CENP-A^{Cnp1} occupancy and kinetochore formation, are even more AT-rich (71.8% [I], 71.9%[II], 70.6% [III]). However, local AT-content, as calculated above, is highly variable across centromeres (Figure 3.1). A comparison between the local AT-content and the sites of CENP-A^{Cnp1} occupancy across all centromeres revealed that the sites occupied by CENP-A^{Cnp1} nucleosomes within all three central domains are less AT-rich (mean=67.0% [sonicated], 67.7% [MNase]) than the gaps separating CENP-A^{Cnp1} occupied sites (mean=72.9% [sonicated], 72.7% [MNase]; Figure 3.2).

Motifs present within the CENP-A^{Cnp1} peaks and gaps were identified using MEME (Bailey and Elkan, 1994). The most significant motif found within the gaps was a highly asymmetrical AT-rich sequence found in 31 out of 51 gaps (Figure 3.3). The sequence features can potentially be analysed in more detail by investigating the frequency distribution of k-mers, which is currently being undertaken (Xu et al., 2012). In contrast, the most significant motif found within the peaks was only present at 9 CENP-A^{Cnp1} peaks (Figure 3.4). This motif contains AA/TT/AT di-nucleotides occurring with 9-11 bp periodicity, which is similar to the described 10.5 bp periodicity of strongly positioned canonical nucleosomes, but this analysis is likely limited by the small number of occupied positions available for analysis. This indicates that *in vivo* CENP-A^{Cnp1} nucleosomes are similar to canonical nucleosomes in that they have a preference against wrapping and occupying highly

A

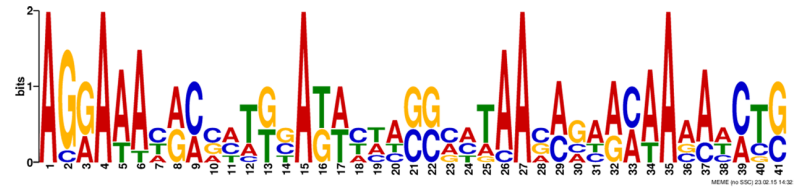


B

Name	Strand	Start	*p*-value	Sites
27	-	486	1.01e-10	TAAAAAAAAA AAAAAAAAAAAAGGAAATTGTAACTTGAA ATTTATGTGA
16	+	390	1.01e-10	AAAAAAAAAA AAAAAAAAAAAAGGAAATTGTAACTTGAA ATTTATGTGA
14	-	284	2.16e-08	GTAGATAATT ATAAACAAAGATATACACAGAAAGTTCAG TTTAAATCCA
2	+	382	2.16e-08	GTAGATAATT ATAAACAAAGATATACACAGAAAGTTCAG TTTAAATCCA
42	+	110	2.69e-08	TATTCAAAAT TAAAAAAAAGAAAGACCTTGAACGATTAA AAATAAATAC
10	+	487	2.69e-08	TATTCAAAAT TAAAAAAAAGAAAGACCTTGAACGATTAA AAATAAATAC
40	-	216	3.35e-08	TCTTTATATG ACAAAAATAAAAACAAATGGCGATTTTAA GCAAAAATAC
9	-	219	3.35e-08	TCTTTATATG ACAAAAATAAAAACAAATGGCGATTTTAA GCAAAAATAC
37	-	458	3.73e-08	TTACCATACG AAAAAATAAAAAGTAAATGATGACTTCAA AAAGATTTGT
6	-	553	3.73e-08	TTACCATACG AAAAAATAAAAAGTAAATGATGACTTCAA AAAGATTTGT
12	+	165	4.62e-08	ATAAAAAAGT ATAAAAGAAAAAAGATTTGGTCACAGCCG GGTTCGAACC
4	-	225	4.62e-08	ATAAAAAAGT ATAAAAGAAAAAAGATTTGGTCACAGCCG GGTTCGAACC
44	+	113	5.69e-08	TGAATGTATT AAAAAAAAATCAATAATTGTAACGATCAG GGGTTTGTAT
32	-	134	5.69e-08	TGAATGTATT AAAAAAAAATCAATAATTGTAACGATCAG GGGTTTGTAT
38	+	383	6.98e-08	CGTAAAATAT TAAAAAAAAAAAATAAAAATAAACTTTCG TTATACATAG
17	+	314	6.98e-08	TACGAAAATC AAAAAAAAAAAGTTGAAGGGTACGTTTAA TTCTTGGTTT
7	+	392	6.98e-08	CGTAAAATAT TAAAAAAAAAAAATAAAAATAAACTTTCG TTATACATAG
13	-	138	7.72e-08	CAGTTTCAAA AAAATAAAAAAAACAGATAGAAATTTGAC TATGCTTTCC
3	+	234	7.72e-08	CAGTTTCAAA AAAATAAAAAAAACAGATAGAAATTTGAC TATGCTTTCC
43	+	473	1.53e-07	CTTATATATG ACACACGAATAACAACATGGCCAGTTGGC CGAGCGGTCT
33	-	68	1.53e-07	CTTATATATG ACACACGAATAACAACATGGCCAGTTGGC CGAGCGGTCT
39	+	163	2.90e-07	AAAGGTTTCA ACGAAAAATAAATAAAACTAAACGTTGAG CTAATTAAAC
8	+	155	2.90e-07	AAAGGTTTCA ACGAAAAATAAATAAAACTAAACGTTGAG CTAATTAAAC
20	-	83	4.11e-07	ACTATCATAG TAAAAAAAAAAAATAAAATAAAAAATCAG TGGTGAAAGC
35	-	148	4.47e-07	TTTGGTTAAA AAAAAAAAAAGAAGTATTGAACAATGGG AAAAATCATA
47	-	132	5.75e-07	CAACTAATCT ACAAAAAAAAAAAAAAGGAATATATTTGG TCATTTAAAC
29	+	619	5.75e-07	CAACTAATCT ACAAAAAAAAAAAAAAGGAATATATTTGG TCATTTAAAC
23	-	185	7.34e-07	CAAGCAAAAC CAAACAGAAAAAAGAAATTTACTGTTTCG CAAATGGTCT
41	+	42	8.61e-07	TCAAGTTTCA AAGAAAAAAAATATATATGGTCTCGTTAA CTTCAGCAGC
31	-	312	1.09e-06	TAAAAAAA AAAAAAAAAAAAACACTTAGTCTCCTTCT ATCAAATAAA
18	+	132	9.77e-06	TCAATTTTTA GAAAAAAAAAAATAACGCATTAATTGTAA ATGTTTTAGG

Figure 3.3. Gaps between CENP-A^{Cnp1} nucleosomes are enriched for asymmetrical A/T rich sequences. The program MEME (Bailey and Elkan 1994) was used to identify motifs enriched within the gaps between CENP-A^{Cnp1} nucleosomes. (A) The most significant motif identified was a 21 bp highly asymmetric A/T rich motif with an E-value of 4.6e-020. (B) This motif was found at 31 of the 51 CENP-A^{Cnp1} gaps.

A



B

Name	Strand	Start	*p*-value	Sites
41	+	66	5.43e-17	ATTTCGCCGA AGGAAACGCCATGGAGTACACGCATAAAACAACAACCACTG AAAGCGAATC
8	+	58	5.43e-17	ATTTCGCCGA AGGAAACGCCATGGAGTACACGCATAAAACAACAACCACTG AAAGCGAATC
48	+	3	9.35e-16	TA AGAAAACACGCTGCAGACATGGGAAAACAGTACAAAATCTG AACATGTTGA
33	-	103	9.35e-16	TTTA AGAAAACACGCTGCAGACATGGGAAAACAGTACAAAATCTG AACATGTTGA
12	-	30	2.26e-13	TTTTGACTTT AGGATAAACCAATTATACTAGCCCTAAGAACGAAAAACCTC AAAAAAAGGC
5	+	53	2.26e-13	TTTTGACTTT AGGATAAACCAATTATACTAGCCCTAAGAACGAAAAACCTC AAAAAAAGGC
40	+	95	2.54e-12	TTAGCAAAAA AGGAAATGAATTTGATTTTCCCATTAACCGAAATAGAAACG TAGAAATAAA
7	+	104	2.54e-12	TTAGCAAAAA AGGAAATGAATTTGATTTTCCCATTAACCGAAATAGAAACG TAG
28	+	99	3.15e-11	TGTACATAAT ACAAATTAAGCCGCATATAAGGACGCAGCGAGCAAAATAGC CTTCCTAA

Figure 3.4. CENP-A^{Cnp1} enrichment peaks share a nucleosome positioning motif. The program MEME (Bailey and Elkan 1994) was used to identify motifs enriched at CENP-A^{Cnp1} peaks. (A) The most significant motif identified was a 41 bp motif with an E-value of 3.8e-014. The motif contains A/T nucleotides occuring with a 9-11 bp periodicity, similar to the described 10.5 bp periodicity of strongly positioned canonical nucleosomes. (B) This motif was found only at 9 of the 54 CENP-A^{Cnp1} peaks.

AT-rich sequences, including asymmetric AT-rich sequences. Consequently CENP-A^{Cnp1} nucleosome positioning within centromeres has a large sequence dependent component. However, it remains possible that other factors also affect this positioning. For example, proteins or protein complexes may bind the AT-rich gaps and consequently position flanking CENP-A^{Cnp1} nucleosomes.

The above analysis show that there is a clear difference between CENP-A occupied sites and intervening gaps. Next the pattern of AT-content was further investigated around the CENP-A peak sites themselves. The peak positions of all CENP-A occupied sites were aligned and the average AT-content was calculated at given distances from the peak positions for 100 base pairs in both directions (Figure 3.2). Comparison of all peaks in this way revealed that the midpoint of CENP-A^{Cnp1} occupancy peaks occurs at the most CG-rich position. Indeed, AT-richness increases in both directions as the distance from the peak position increased, with even small distances from the peak having a higher average AT-content. Thus, CENP-A^{Cnp1} nucleosomes are generally positioned at a single most favourable site with even nearby flanking sequences being less favoured. This pattern may reflect an evolved feature of fission yeast centromeres related to factors involved in the deposition of CENP-A^{Cnp1}, CENP-A^{Cnp1} nucleosomes themselves, or its function as the site of kinetochore attachments, among other possibilities.

3.2.2. Occupancy algorithm predicts additional CENP-A^{Cnp1} occupied sites within the central domain

Previous analysis (Miell and Berger, Allshire Lab) found that CENP-A^{Cnp1} occupancy within the central domains could be predicted with moderate accuracy. This suggested that specific sequence features influenced CENP-A^{Cnp1} occupancy within the central domains. At the same time, this also suggested that sequence features alone did not fully explain the pattern of CENP-A^{Cnp1} occupancy. The CENP-A^{Cnp1} ChIP-seq data of Lando et al. (2012) was re-evaluated to determine what aspects of CENP-A^{Cnp1} occupancy could and could not be predicted based on sequence features.

Genome-wide *S. pombe* H3 nucleosome occupancy and positioning predictions were downloaded from the Segal lab (http://genie.weizmann.ac.il/software/nucleo_genomes.html). These values were

calculated using the published nucleosome-DNA interaction model (Kaplan et al., 2009; Segal et al., 2006). H3 nucleosome occupancy predictions were used since there was a high correlation between H3 and CENP-A^{Cnp1} occupancy *in vitro* and also between the predictive power of H3 and CENP-A^{Cnp1} occupancy models. Genome-wide occupancy predictions were compared to the CENP-A^{Cnp1} ChIP-seq data from sonicated chromatin.

There is a general decrease in the probability of nucleosome occupying sites within the central domain relative to the rest of the genome, likely due to the higher AT-content of the central domain. Nucleosomes are predicted to occupy discretely positioned sites within the central domains (Figure 3.5). Within the central domains there was a moderate correlation between predicted nucleosome occupancy and observed CENP-A^{Cnp1} occupancy (τ =0.380; Figure 3.6A). Most sites show a positive correlation between predicted and observed occupancy, as expected. However, some sites predict the presence of an occupying nucleosome, but have low enrichment of CENP-A^{Cnp1} *in vivo*. These unoccupied sites do not reflect scattered noise, for example in the case that the positioning of *in vivo* and predicted sites was slightly misaligned. Rather, a second class of positioned nucleosomes is predicted to occupy the gap regions between the CENP-A^{Cnp1} nucleosome positions (Figure 3.5). These sites may be occupied by CENP-A^{Cnp1} *in vivo*, but CENP-A^{Cnp1} at these sites could be unstable or transient and therefore could not be detected. Alternately, these sites may be occupied by a different protein complex or may even exist as naked DNA.

At 10,000 random sites genome-wide the predicted nucleosome occupancy varies across a large range of values that are consistent with sites that are predicted to be nucleosome depleted, sites that are strongly predicted to be occupied, and sites with intermediate predicted occupancy (Figure 3.6B). CENP-A^{Cnp1} occupancy is centromere specific and consistent with this there is low CENP-A^{Cnp1} IP/IN coverage at the genome-wide sites. However, CENP-A^{Cnp1} coverage is especially low at sites that are predicted to be nucleosome depleted, suggesting that there is a bias against extracting or sequencing these regions.

The moderate predictive accuracy of the nucleosome occupancy models does not arise form an inability to predict occupied sites accurately. The model successfully

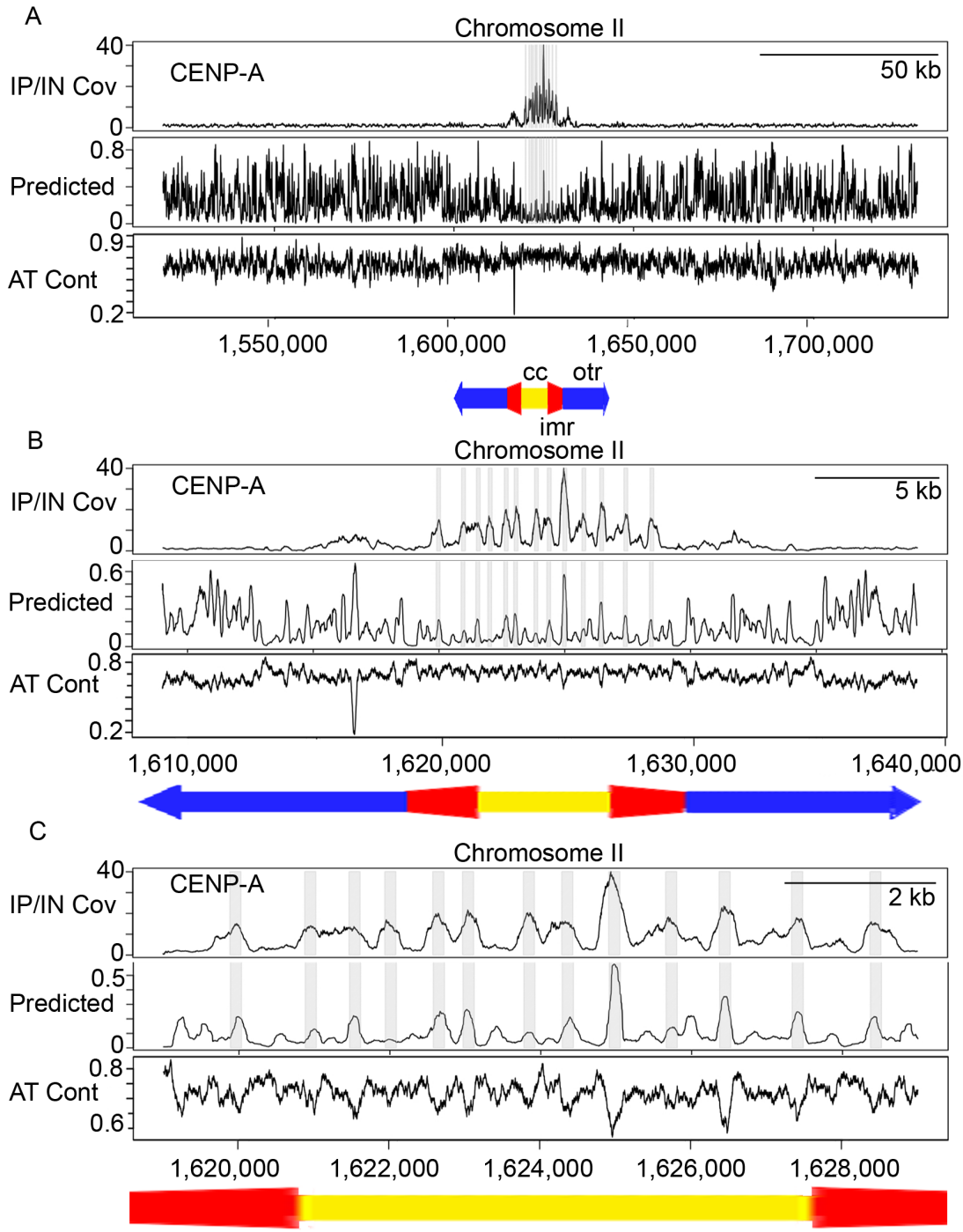


Figure 3.5. Accuracy of nucleosome prediction models. CENP-A^{Cnp1} ChIP-seq using sonicated sheared chromatin (Lando et al. 2012) and predicted nucleosomes occupancy (Segal et al. 2006) are visualised at the central domain of the chromosome II centromere with (A) 100 kb, (B) 10 kb, and (C) 0 kb of flanking sequences. Predicted nucleosome occupied sites correspond to observed CENP-A^{Cnp1} occupied sites, but there are additional predicted sites between CENP-A^{Cnp1} nucleosomes.

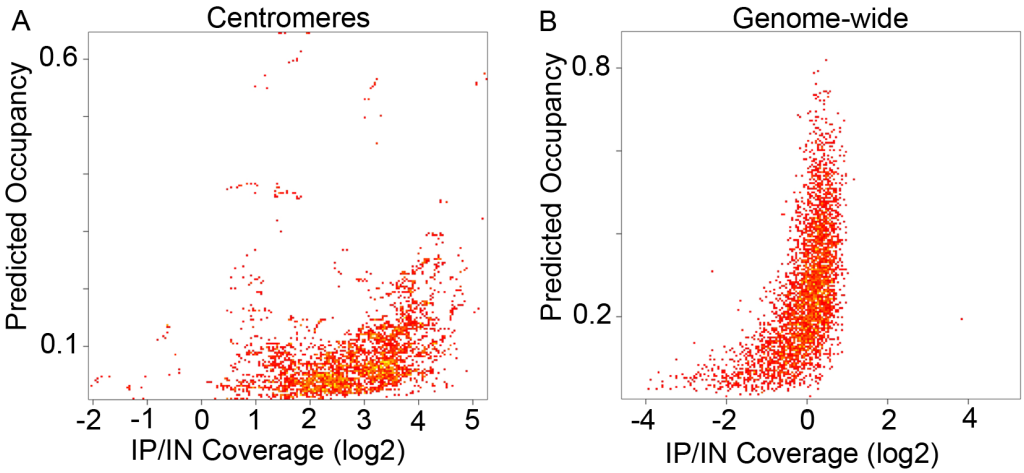


Figure 3.6. Observed CENP-A^{Cnp1} occupancy partially agrees with predicted occupancy. (A) The correlation between predicted nucleosome occupancy and CENP-A^{Cnp1} occupancy shows that there is a positive correlation (τ =0.380), but there is a second class of predicted sites that are not occupied in vivo. (B) This pattern is not observed genome-wide, since there are no regions of CENP-A^{Cnp1} enrichment outside of the centromeres.

predicts the CENP-A^{Cnp1} occupied sites within centromeres based on the underlying sequence features. The primary shortcoming of the models is their inability to predict the second class of unoccupied sites in the gaps between *in vivo* verified CENP-A^{Cnp1} nucleosomes. The lack of occupancy at these sites is therefore not due to sequence features alone.

3.2.3 Published data potentially over-estimates the positioning of CENP-A^{Cnp1} nucleosomes

The above analyses indicate that CENP-A^{Cnp1} preferentially occupies less AT-rich sequences. However, high throughput sequencing technology has a known bias against AT-rich sequences (Benjamini and Speed, 2012; Meyer and Liu, 2014). The bias in ChIP-seq data towards CG-rich sequences poses a potential confounding factor in determining spatial occupancy at the required resolution, especially given that the CENP-A^{Cnp1} occupied regions of fission yeast centromeres are comprised of AT-rich sequences. Thus, it was necessary to evaluate the level of bias present in the published CENP-A^{Cnp1} ChIP-seq data.

There is the high correlation between the sonicated and MNase CENP-A^{Cnp1} ChIP-seq data sets within the central domains, but an analysis of the sequenced sheared input chromatin (IN) reveals that there is evidence of bias in the data. A random sampling of 10,000 single nucleotides across the genome (excluding centromeres and regions proximal to telomeres) shows that the data is subject to the previously described bias, leading to a negative correlation between input (IN) coverage and AT-content for both the sonicated (τ= -0.467) and MNase digested data (τ =-0.561; Figure 3.7A/B). This bias becomes more prominent at high AT-content sequences. This bias is more significant for the MNase digested data. The extensive bias towards higher coverage of CG-rich sequences in the MNase data is at least partially explained by the fact that nucleosomes preferentially occupy CG-rich sites and will therefore be more resistant to MNase digestion.

The reduced extraction efficiency of centromere regions (i.e. the lower recovery of these regions in the IN) means that direct comparisons between centromeres and the rest of the genome cannot provide information on the sequencing bias specifically at centromeres. However, examination of the IN coverage within the central domains centromeres will determine if the same sequencing bias occurs

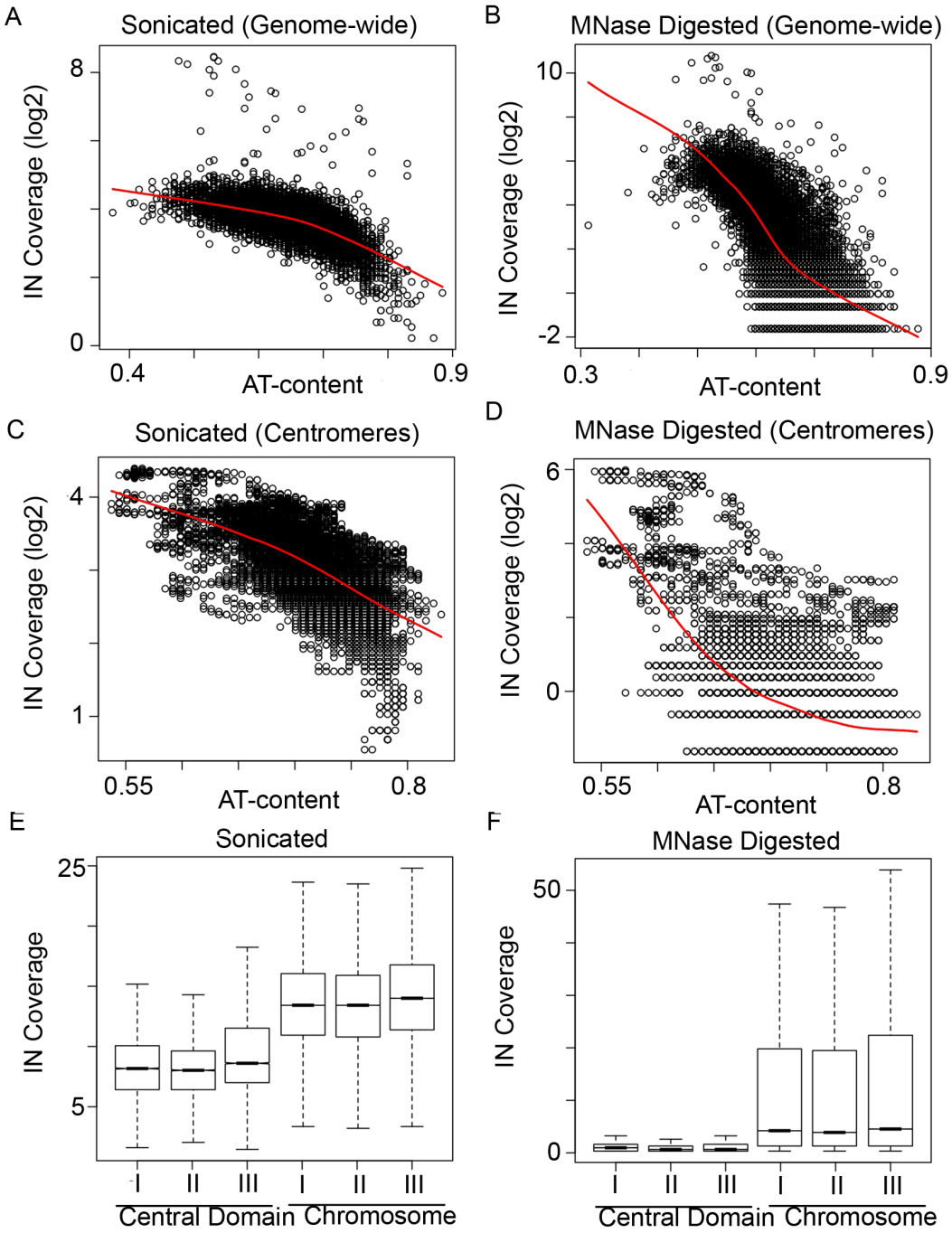


Figure 3.7. Published data is affected by sequencing bias. The IN sonicated and MNase digested whole cell extract data showed evidence of bias in extraction and Illumina sequencing. At 10,000 randomly chosen sites across the genome AT-content was negatively correlated with IN coverage for (A) sonicated (τ= -0.467) and (B) MNase digested data (τ =-0.561). Similarly, at 10,000 sites within the central domain there was a negative correlation between AT-content an IN coverage for (C) sonicated (τ=-0.461) and (D) MNase digested data (τ=-0.318). Additionally, there is evidence of extraction bias of the centromere with an ~50% drop in centromere coverage relative to the rest of the genome and extremely low IN coverage of MNase digested data.

within these regions of interest. 10,000 sites were randomly chosen from within the three central domains and correlated with local AT-content. Again a negative correlation between IN coverage and AT-content within the central domains for both the sonicated (τ=-0.461) and MNase digested data (τ =-0.318; Figure 3.7C/D) was observed.

Both datasets have lower coverage of all three centromere central domains, with an ~50% drop in coverage in the sonicated data and extremely low coverage in the MNase data (Figure 3.7E/F). This under-representation has been observed before as measured by qRT-PCR (A. Pidoux, Allshire lab, personal communication), and may reflect an extraction bias due to the assembly of kinetochore complexes. In addition, the extremely low IN coverage for the MNase digested data indicated that the sample might have been over-digested.

Initial evidence indicates that the bias likely affects sequence coverage and thus enrichment at CENP-A^{Cnp1} peaks. Both the sonicated (Figure 3.8) and MNase digested (Figure 3.9) data sets display uneven coverage across CENP-A^{Cnp1} occupied regions within the centromere. Within the chromosome II centromere this is characterised by a single primary peak with high coverage and other peaks with more moderate levels of coverage. This enrichment towards a single primary peak is even more apparent in the MNase digested data. This single peak is associated with the most CG-rich region within the centromere and thus sequencing bias may lead to this exaggerated variation in peak enrichment. It remains possible that this pattern is due to variable levels of CENP-A^{Cnp1} occupancy across the centromere, but the greater magnitude of the difference in the MNase data in particular supports a contribution of sequencing bias to this pattern. The heterochromatic outer-repeats show higher coverage in the IN (Figure 3.8), but this is likely due to an underestimation of the number of repeats in the genome assembly. Sequencing bias must be taken into account before an accurate and unbiased analysis of CENP-A^{Cnp1} positioning can be undertaken.

3.2.4: Current methodologies allow more accurate and less biased mapping of proteins within the centromeres

The re-analysis of published ChIP-seq data indicates that the methodology has a bias towards CG-rich sequences, which may affect the reliability of the data and

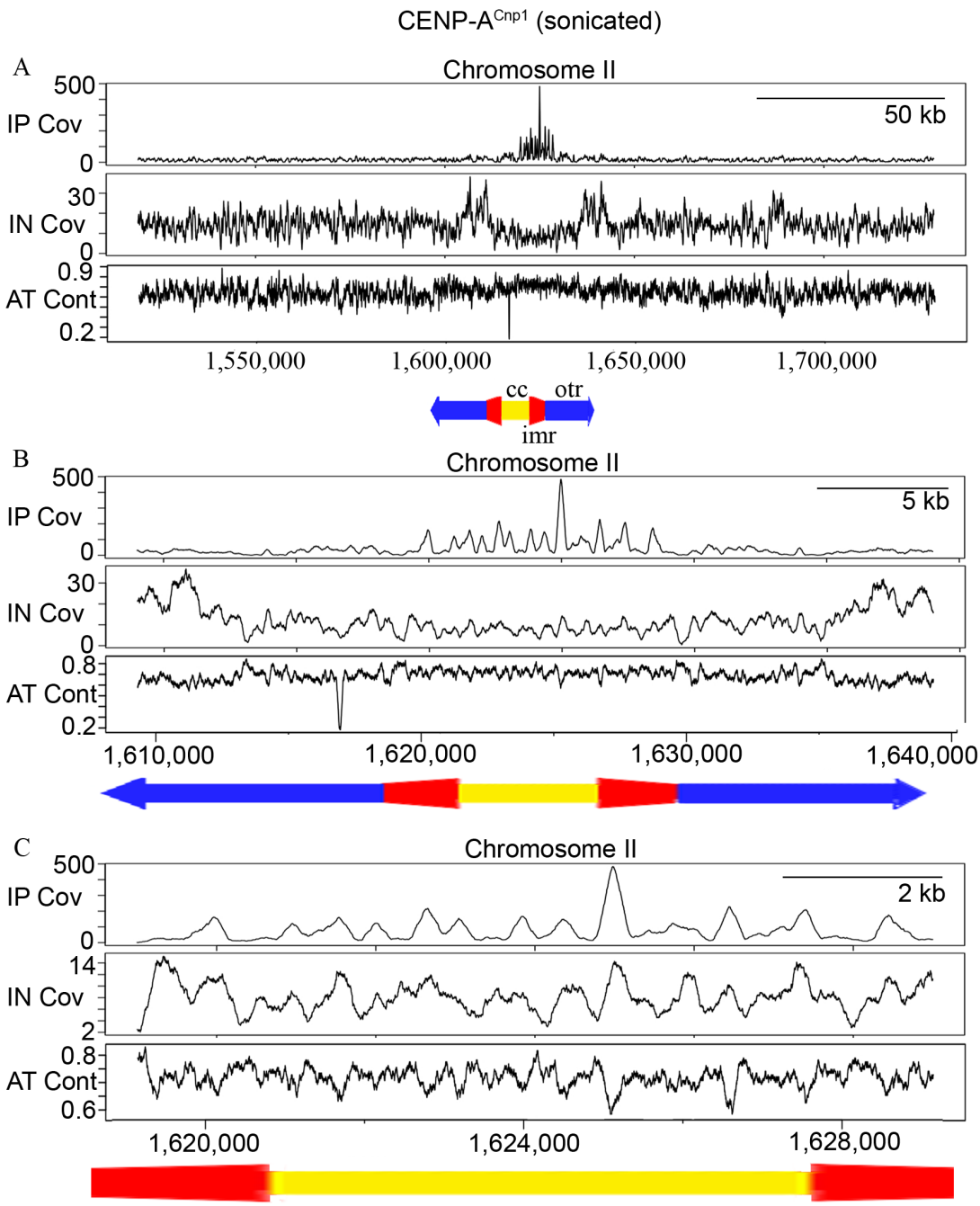


Figure 3.8. Sequencing bias affects ChIP-seq coverage. Fragment coverage at the chromosome II centromere is shown, including (A) 100 kb, (B) 10 kb, and (C) 0 kb of flanking sequences around the central domain for CENP-A^{Cnp1} ChIP-seq IP and IN. When this is compared to the local AT-content it can be seen that the single highest enrichment peak corresponds to the region with lowest AT-content. Higher IN coverage is also associated with this peak and other regions of lower AT-content. There is an apparent higher IN coverage of the centromeric outer repeats, but this likely reflects an underestimation of the number of repeats present.

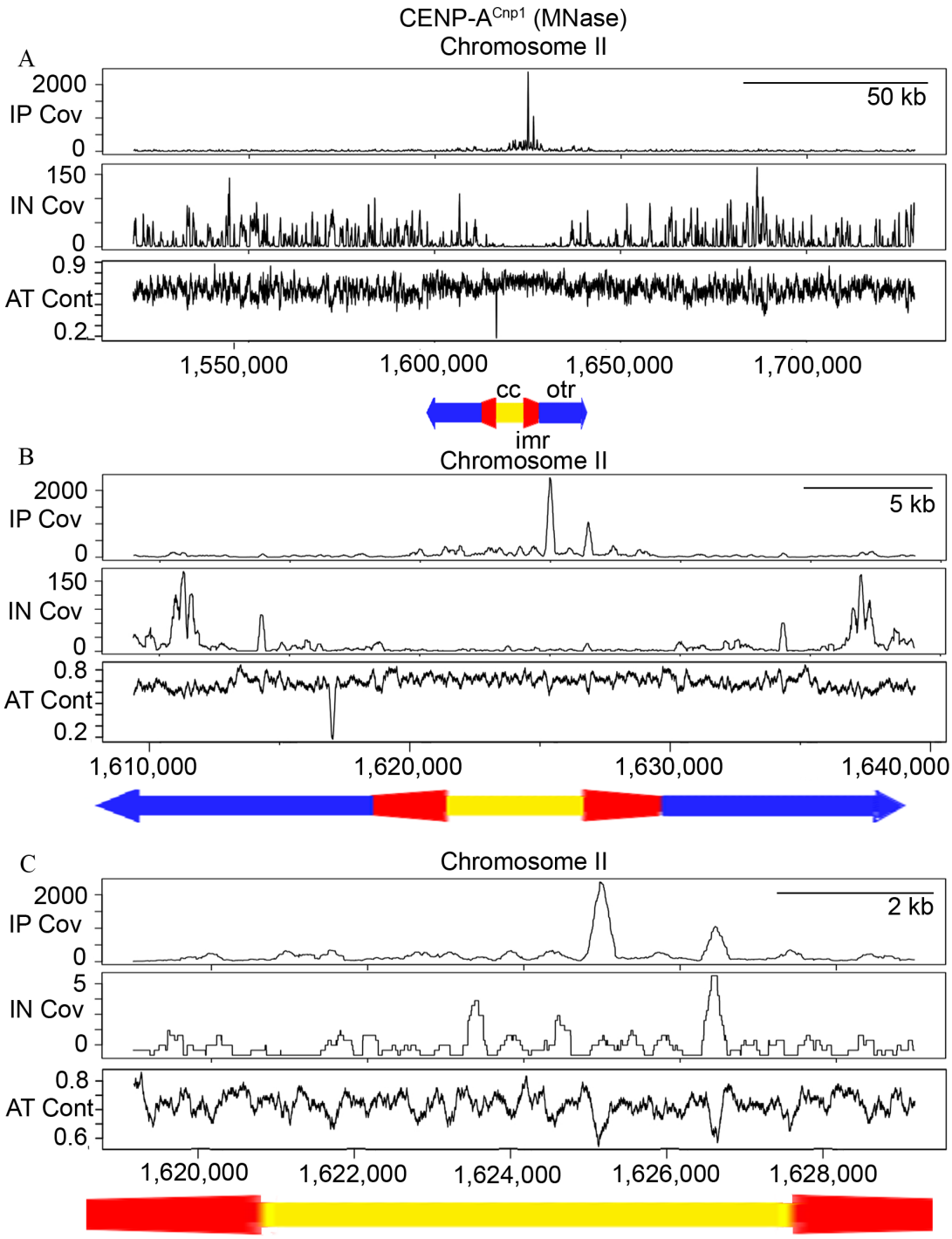


Figure 3.9. Sequencing and MNase digestion bias affects ChIP-seq data. Fragment coverage at the chromosome II centromere is shown, including (A) 100 kb, (B) 10 kb, and (C) 0 kb of flanking sequences around the central domain for MNase digested CENP-A^{Cnp1} ChIP-seq IP and IN. Similarly to the sonicated data, the most enriched peak corresponds to the region with the lowest AT-content and the second most enriched peak corresponds to the region with the second lowest AT-content. This effect is more substantial than observed in the sonicated data and likely reflects biases in MNase digestion and Illumina sequencing. Substantial underrepresentation of the central domain in the IN suggests an issue with centromere extraction or nucleosome over-digestion.

analyses of sequence features affecting CENP-A^{Cnp1} positioning. The technology has continued to advance and it was necessary to re-evaluate the biases in sequencing technologies to determine whether current methodologies were more reliable for high resolution mapping of centromere and kinetochore components.

CENP-A^{Cnp1} ChIP-seq was thus performed using the same protocol as originally used for CENP-A^{Cnp1} ChIP-seq data (Lando et al., 2012). Anti-CENP-A^{Cnp1} ChIP was carried out on wild type cell extracts. In addition anti-GFP ChIP was performed on extracts from cells expressing N-terminally tagged GFP-CENP-A^{Cnp1}. Chromatin sheared by sonication was used to avoid the inherent biases associated with MNase digestion. Sheared chromatin from purified whole cell extract (IN) was sequenced to determine the sequencing and extraction biases present in current sequencing methodologies and technologies. Initial Illumina library making attempts were unsuccessful and unreliable due to the limited amount of DNA obtained from ChIP experiments. Therefore, the Illumina library making protocol was adapted for use with very low amounts of DNA (see chapter 2) resulting in a method that was very reliable down to at least a minimum of 0.25 ng of starting DNA. Libraries were prepared using 5 ng of input DNA. Libraries were subjected to 100 bp paired end sequencing.

Purified sonicated chromatin (IN) was distributed with a mode of ~300 bp and a large proportion of fragments in the size range of interest (100-200 bp) allowing CENP-A^{Cnp1} positions to confidently mapped (Figure 3.10). Illumina libraries consistently were tightly distributed around ~300 bp, which reflected a fragment size of ~150 bp after subtracting length of the adapters. Additional post hoc size selection could be done since paired end sequencing gave absolute information of the size of all fragments. Following 100 bp paired end sequencing, assessment by FASTQC revealed that the data was extremely high quality. Quality scores measured by the machine were consistently high (Phred score > 30) across reads within the sample and remained high over the entire read length. The GC-content of reads were unimodally distributed around ~34%, consistent with the genome-wide GC-content of *S. pombe*. GC-content was also uniform across the length of reads with very few over-represented sequences, indicating that contamination from library primers or other sources was not significant.

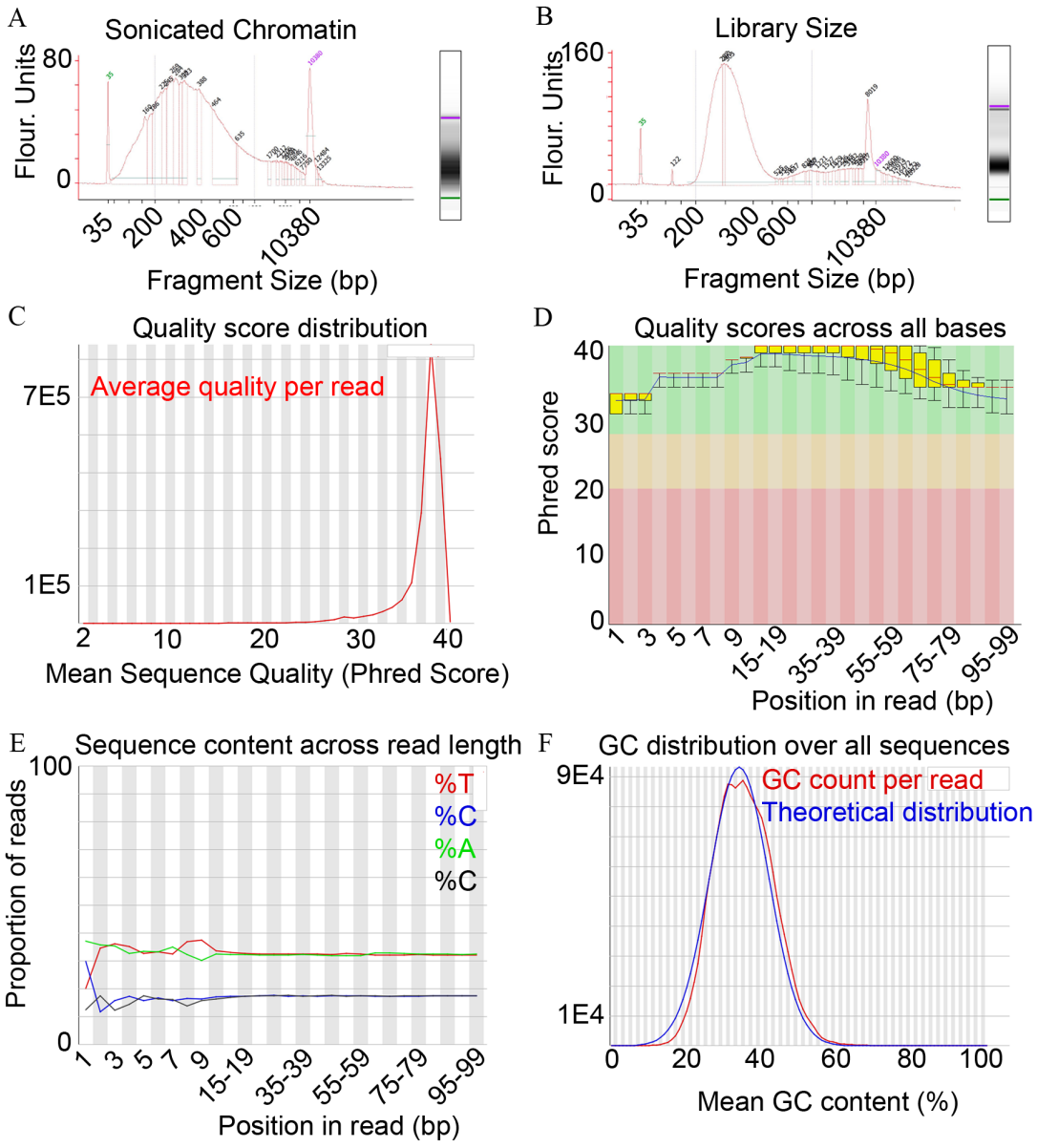


Figure 3.10. New CENP-A^{Cnp1} ChIP-seq is high quality. (A) Chromatin shearing and purification resulted in a pool of fragments distriburted around a mode of ~300 bp with a large proporiton of fragments in the range of interest (100-250 bp). (B) Library making of ChIPed DNA resulted in libraries with tight size selection distributed around ~300 bp, corresponding to a sequenced fragment size of ~150 bp. (C) Illumina sequencing resulted in reads with high quality, with most reads hacing an average Phred score >30. (D) Across the length of reads the quality remained high. (E) Nucleotide content is uniform across the length of reads. (F) Across reads the nucleotide content is unimodally distributed with values consistent with data from *S. pombe*.

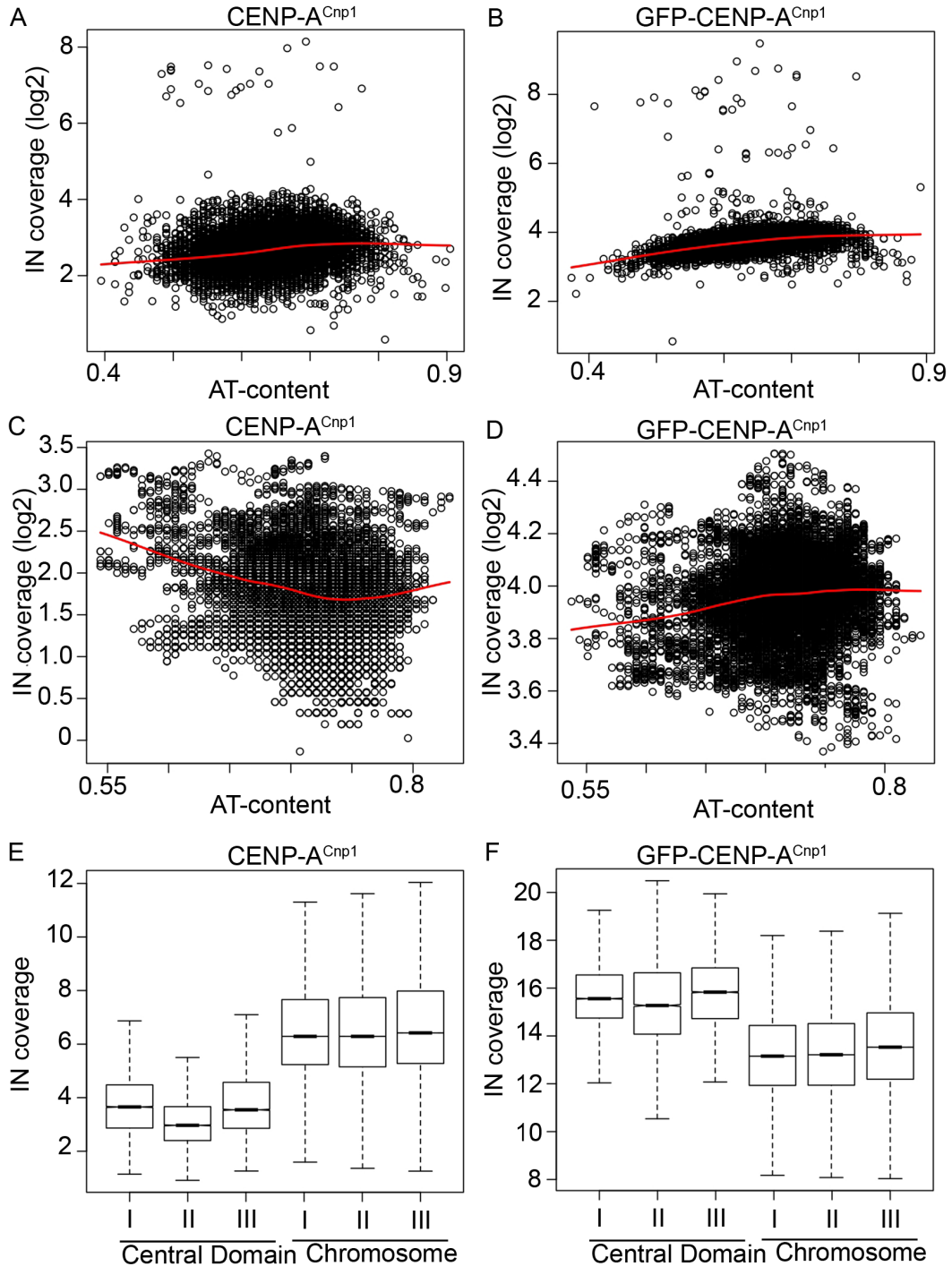


Figure 3.11. New CENP-A^{Cnp1} ChIP-seq data has less sequencing bias. (A) IN coverage at 10,000 random sites genome-wide does not show a negative correlation between coverage and AT-content for (A) wild type (τ=0.165) or (B) GFP-CENP-A^{Cnp1} strains (τ=0.392). Within the central domain there is no correlation between AT-content and IN coverage for the (C) wild-type strain (τ=-0.134) or the (D) GFP-CENP-A strain (τ=0.072). Central domain IN coverage is lower than the chromosomal averages for the wild type strain (E), but slightly higher for the GFP-CENP-A^{Cnp1} strain (F).

The IN samples were analysed to determine whether the previously described sequencing bias was present in this new data set. There was no correlation between AT-content and IN coverage for wild-type strain (τ=0.165) and a weak positive correlation for the GFP-CENP-A^{Cnp1} strain (τ=0.392; Figure 3.11) at 10,000 random sites across the genome. This is in contrast to the moderate negative correlation observed with the published data (τ=-0.467) from Lando et al. (2012). Within the central domain of centromeres there is no correlation between AT-content and IN coverage for the wild-type strain (τ=-0.134) or the GFP-CENP-A^{Cnp1} strain (τ=0.072). This is in contrast to the moderate negative correlation observed for the published data (τ=-0.461) from Lando et al. (2012). The wild type IN data had lower coverage of centromeres, while centromere sequences were not underrepresented in the GFP-CENP-A^{Cnp1} IN data. This indicates that the coverage issue with centromere regions does not result from a technical bias in high throughput sequencing methodology. An extraction bias from cells is a possible explanation for the lower centromere coverage. It is unclear what could lead to the difference in extraction of the central domain in the strain with GFP-CENP-A^{Cnp1}. It is possible that the GFP tag on CENP-A^{Cnp1} alters the structure of kinetochore attachments without altering viability. Nevertheless, the data are high quality and can be used to obtain a more accurate map of CENP-A^{Cnp1} occupancy at centromeres.

3.2.5. New CENP-A^{Cnp1} ChIP-seq data suggests fuller occupancy and supports positioning based on sequence content

The newly sequenced CENP-A^{Cnp1} ChIP-seq IP and IN data are high quality, both in terms of high sequencing quality and low AT-content bias. The data will provide a more accurate representation of CENP-A^{Cnp1} positioning and occupancy. The analyses of CENP-A^{Cnp1} occupancy and positioning applied above to the Lando et al. (2012) published data were repeated on this new CENP-A^{Cnp1} and GFP-CENP-A^{Cnp1} ChIP-seq sequence data.

Focusing on the chromosome II centromere, the wild-type CENP-A^{Cnp1} and GFP-CENP-A^{Cnp1} ChIP-seq data show that CENP-A^{Cnp1} is highly enriched across the central domain (Figure 3.12). This new wild-type CENP-A^{Cnp1} ChIP-seq data suffers from an extraction bias in the IN sample that results in low coverage of centromere DNA and a high level of noise if this is used for normalisation. For this reason the IP

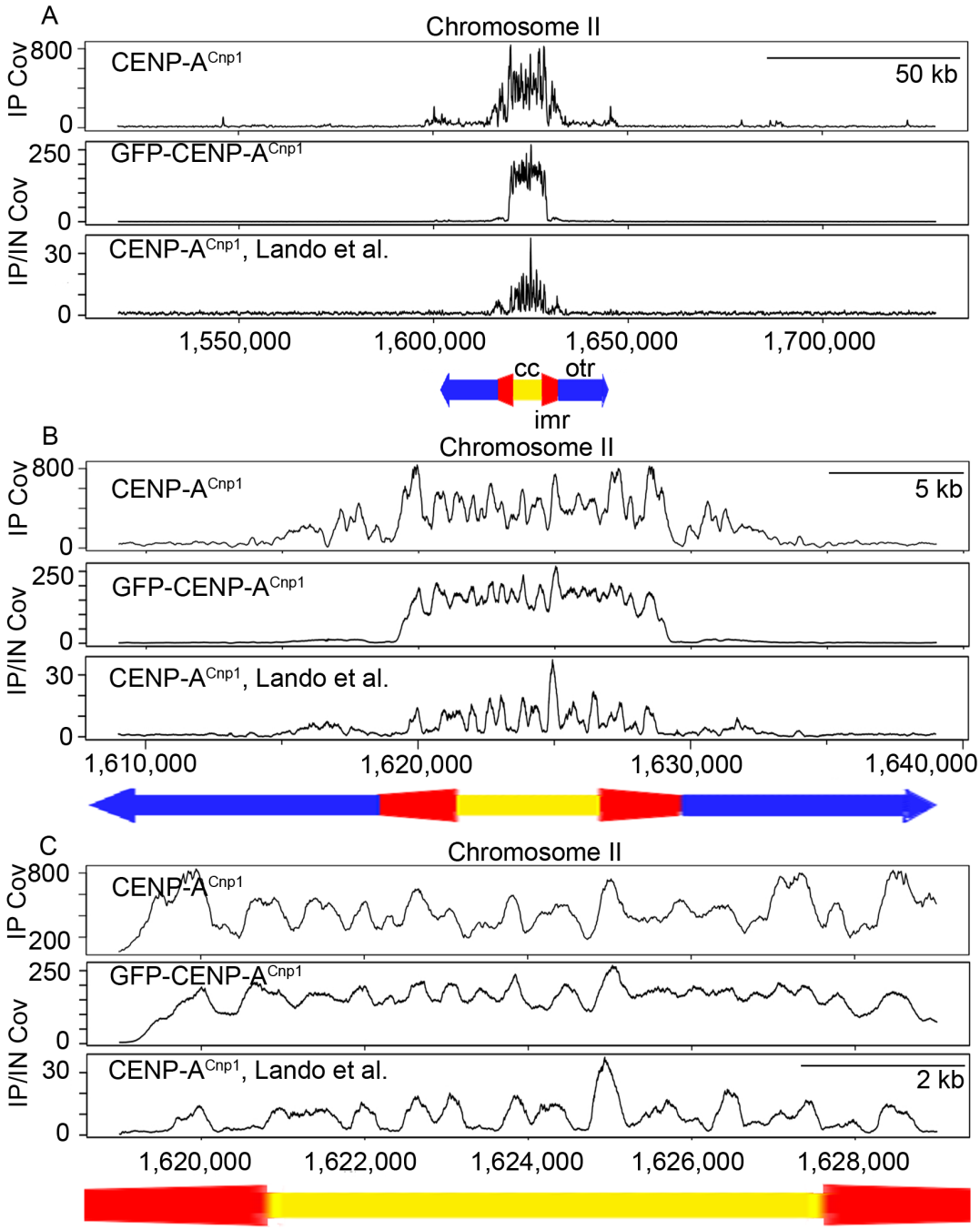


Figure 3.12. New CENP-A^{Cnp1} ChIP-Seq shows that CENP-A^{Cnp1} is enriched at centromeres. (A) ChIP-seq was performed using anti-CENP-A^{Cnp1} serum in a wild-type strain and anti-GFP antibody in a GFP-CENP-A^{Cnp1} strain. This was compared to published CENP-A^{Cnp1} ChIP-Seq data (Lando et al. 2012) and visualised at the chromosome II centromere, including (A) 100 kb, (B) 10 kb, and (C) 0 kb flanking sequences. The new CENP-A and GFP-CENP-A^{Cnp1} ChIP-seq shows that CENP-A^{Cnp1} is highly enriched at the centromere, but GFP-CENP-A^{Cnp1} is highly enriched across the entire central domain.

coverage (normalised by millions of reads mapped) was used for analyses instead. There is a low level of enrichment over the outer-repeats (indicated in blue), but this enrichment is greatly reduced when the IP/IN ratio is used, suggesting that the enrichment is driven by higher coverage of the outer-repeats in both the IP and IN. This is likely due to an underestimation of the number of outer-repeat sequences present in the genome.

There is a strong correlation between IP coverage of the previously published CENP-A^{Cnp1} data and the newly sequenced wild-type αCENP-A^{Cnp1} data sets within the central domains (τ=0.73; Figure 3.13A). However, there are also clear differences between the two data sets. The new wild-type CENP-A^{Cnp1} ChIP-seq data has high enrichment over the entire central domain of each centromere (only the chromosome II centromere is shown) with the peaks enriched to an even greater extent, while the previously published data had enrichment specific to the defined CENP-A^{Cnp1} peaks with very low coverage in the gaps between peaks (Figure 3.12). The high overall level of enrichment makes peak calling more difficult. It is unclear whether the high level of enrichment across the entire central domain is due to actual CENP-A^{Cnp1} occupancy, cross-linking to flanking DNA, or even cross-linking to another protein complex. Nevertheless, the newly sequenced CENP-A^{Cnp1} data supports the observation that CENP-A^{Cnp1} is highly positioned at specific sites on more CG-rich sequences (68.7% AT-content) with AT-rich gaps (72.8%; $p<0.001$) and a median gap size of 532.5 bp (Figure 3.14), although the gap size is likely an overestimation due to difficulty in calling peaks.

The N-terminally tagged GFP-CENP-A^{Cnp1} data shows a greater difference in comparison with the other data sets (Figure 3.12). The GFP-CENP-A^{Cnp1} data is still moderately correlated with the CENP-A^{Cnp1} ChIP-seq data of Lando et al. 2012 (τ=0.500; Figure 3.13.B) and the new wild-type CENP-A^{Cnp1} ChIP-seq data (τ=0.497; Figure 3.13C). However, the GFP-CENP-A^{Cnp1} ChIP-seq data shows broad occupancy across the entire central domains with only a small additional enrichment at the previously defined peaks, making subpeak calling within the central domain impossible. The CENP-A^{Cnp1} in this strain is not wild type as it is GFP-tagged, though the tag does not affect growth or viability, and the difference is likely due the presence of a large GFP protein tag on all endogenous CENP-A^{Cnp1} (Lando et al., 2012). This is not due to an inherent limitation of ChIP-seq using an anti-GFP

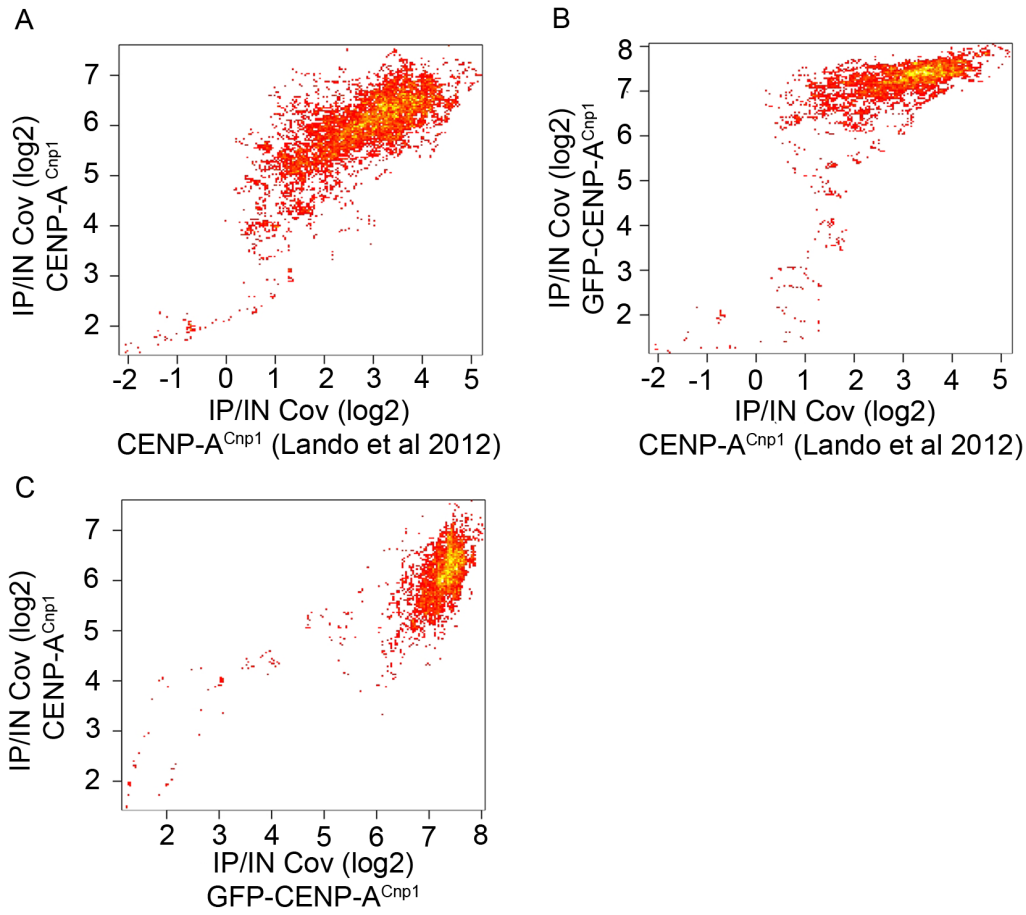


Figure 3.13. New CENP-A^{Cnp1} ChIP-seq is in agreement with published data. (A) At the central domains the three data sets are highly correlated, especially the new and published CENP-A^{Cnp1} ChIP-seq data (τ=0.528). (B) The GFP-CENP-A^{Cnp1} data also correlates with the published CENP-A^{Cnp1} data (τ =0.500) and (C) new CENP-A^{Cnp1} ChIP-seq data (τ=0.497), but the GFP-CENP-A^{Cnp1} is restricted to a smaller range of values, consistent with its more uniform enrichment.

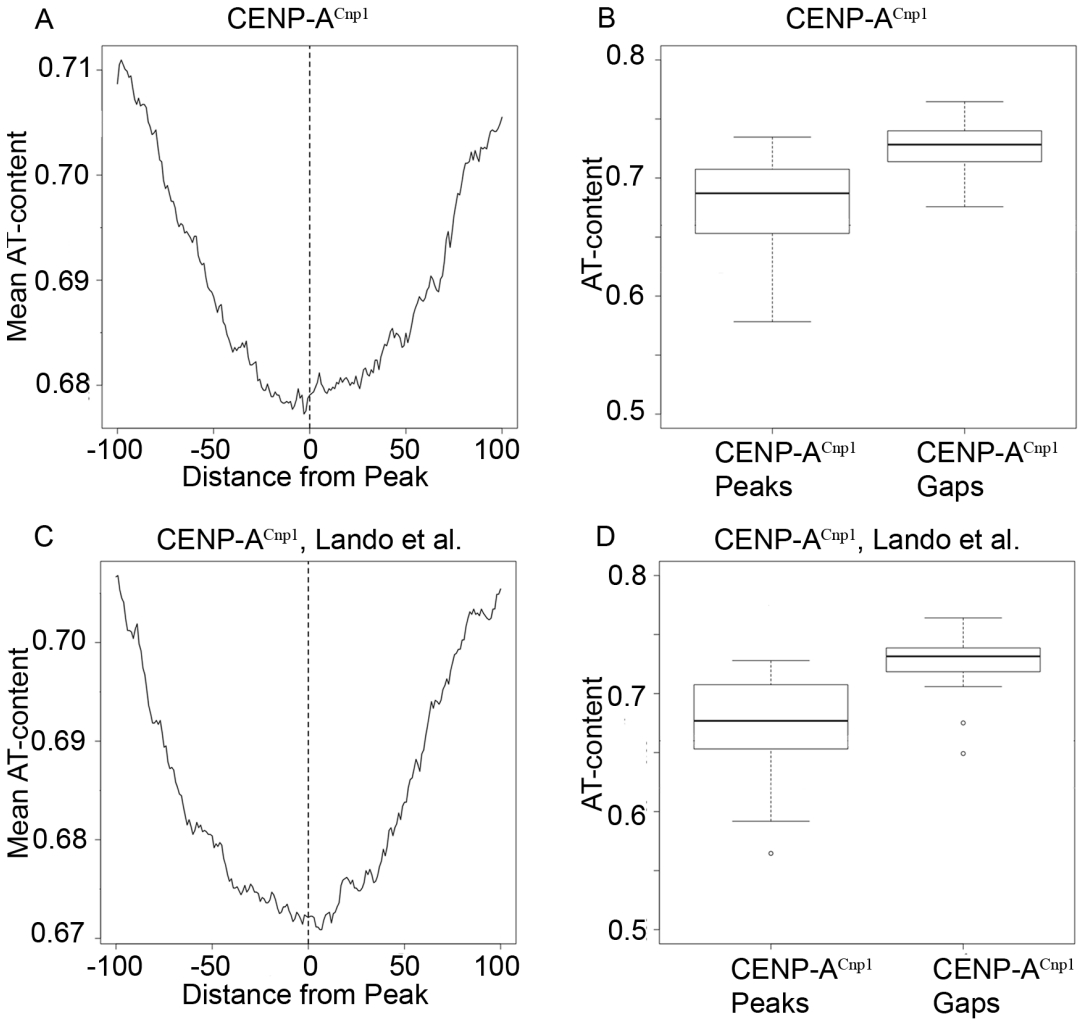


Figure 3.14. New CENP-A^{Cnp1} ChIP-seq supports sequence dependent positioning. (A) CENP-A^{Cnp1} ChIP-seq indicates that CENP-A^{Cnp1} peaks are positioned at sites of minimal AT-content with increasing AT-content at even nearby sites. (B) The CENP-A^{Cnp1} peaks occupy sites with lower AT-content (68.7%) and are separated by gaps with higher AT-content (72.8%; $p<0.001$). (C) This is similar to the published CENP-A^{Cnp1} ChIP-seq data, which shows CENP-A peaks occupying a local AT-content minimum with (D) low AT-content peaks (67.0%) separated by gaps with higher AT-content (72.9%; $p<0.001$).

antibody since other centromeric and kinetochore proteins do not show this pattern (Chapter 5). The altered pattern may be due to disruption in CENP-A^{Cnp1} positioning due to the presence of the GFP tag or by an increased degree of protein-protein or protein-DNA crosslinks caused by the large GFP tag (double the size of CENP-A^{Cnp1} itself). This could result in some cross-linking of the GFP tag with DNA in the gaps between bona fide CENP-A^{Cnp1} nucleosomes. It would be ideal to repeat this experiment using a smaller tag on CENP-A^{Cnp1}.

The more uniform pattern of high enrichment of GFP-CENP-A^{Cnp1} across centromeres is nonetheless an important observation. Although bias is reduced in the new data sets, the lowest IP coverage within the centromere is associated with the highest AT-rich sequences in the wild-type CENP-A^{Cnp1} ChIP-seq data. No protein has been observed to associate by ChIP to these gaps so it remained unclear whether this represented the true profile of CENP-A^{Cnp1} or whether this was due to technical limitations. Additionally it remained unclear whether it was even possible to ChIP a protein to these regions. Thus the high level of enrichment of GFP-CENP-A^{Cnp1} at all positions across the three central domains indicates that the gaps between CENP-A^{Cnp1} nucleosomes do not reflect a technical artefact. Further experiments could potentially identify gap-binding proteins using these methods. Still, the GFP-CENP-A^{Cnp1} most likely does not reflect the wild-type occupancy profile of CENP-A^{Cnp1} and will not be discussed further.

3.2.6 Less discrete enrichment of CENP-A^{Cnp1} in new ChIP-seq is not due to technical limitations

The new CENP-A^{Cnp1} ChIP-seq data generated has less well-defined peaks of enrichment within the central domains and overall has a high level of enrichment across the entire central domain of all three centromeres. This data is less influenced by technical bias, but it is unclear whether the results are in fact more accurate. This CENP-A^{Cnp1} ChIP-seq data may suffer from higher noise and lower enrichment due to the use of agarose beads. The data in Lando et al. (2012) was also produced using agarose beads but the sequencing bias against AT-rich sequences may have masked this non-specific enrichment. Magnetic Dynabeads can be used to obtain higher enrichment and lower background. This methodology could not be used to for CENP-A^{Cnp1} ChIP-seq because ChIP using the anti-CENP-A^{Cnp1} antiserum with Dynabeads leads to low enrichment for unknown reasons.

ChIP-seq of the inner-kinetochore component CENP-T^{Cnp20} using agarose beads and Magnetic Dynabeads were compared to determine whether CENP-A^{Cnp1} data is influenced by technical limitations due to the ChIP methodology.

ChIP-seq of GFP-tagged CENP-T^{Cnp20} using both agarose beads and Dynabeads resulted in high enrichment specific to centromere central domains (Figure 3.15). As expected, ChIP-seq using Dynabeads resulted in higher enrichment and reduced noise. Neither data set showed the high level enrichment across the entire central domain that was observed for the new CENP-A^{Cnp1} ChIP-seq data. Both datasets have highly positioned peaks within the central domain with low coverage between peaks. The data obtained using agarose beads showed some additional secondary peaks in the gaps between the CENP-T^{Cnp20} peaks that are not observed in the Dynabead data. This was unexpected given that agarose beads have a higher background and lower enrichment and it is not clear which represents the true pattern of CENP-T^{Cnp20}. ChIP-exo (Rhee and Pugh, 2011) will allow even more accurate mapping of protein binding and future application of this methodology will help to resolve the uncertainty in mapping CENP-T^{Cnp20} and CENP-A^{Cnp1}.

Different biases in the two ChIP protocols (using agarose beads or Dynabeads) could cause the differences in enrichment within the central domains. Therefore, two replicates of mock IP ChIP-seq were performed in a wild type strain with no tagged proteins using anti-GFP antibody and either agarose beads or Dynabeads to determine the technical bias of each ChIP methodology. AT-content was negatively correlated with agarose bead IP coverage (τ=-0.523; Figure 3.16A) and Dynabead IP coverage (τ=-0.596; Figure 13.17A) genome-wide. In contrast, AT-content was weakly positively correlated with Agarose bead IN coverage (τ=0.365; Figure 3.16B) and uncorrelated with Dynabead IN coverage (τ=0.101; Figure 3.16B) at 10,000 random sites across the genome. Similarly, AT content was negatively correlated with agarose bead IP coverage (τ=-0.560) and Dynabead IP coverage (τ=-0.469) within the central domain. At 10,000 random sites within the central domain AT-content is uncorrelated with both agarose bead IN coverage (τ=0.009) and Dynabead IN coverage (τ=-0.159). These results suggest that there is no sequencing bias against AT-rich sequences, but that the ChIP protocol itself introduces some bias. This is unlikely to affect the overall observations since, as the GFP-CENP-A^{Cnp1} ChIP-seq showed, high enrichment at all sites within the central

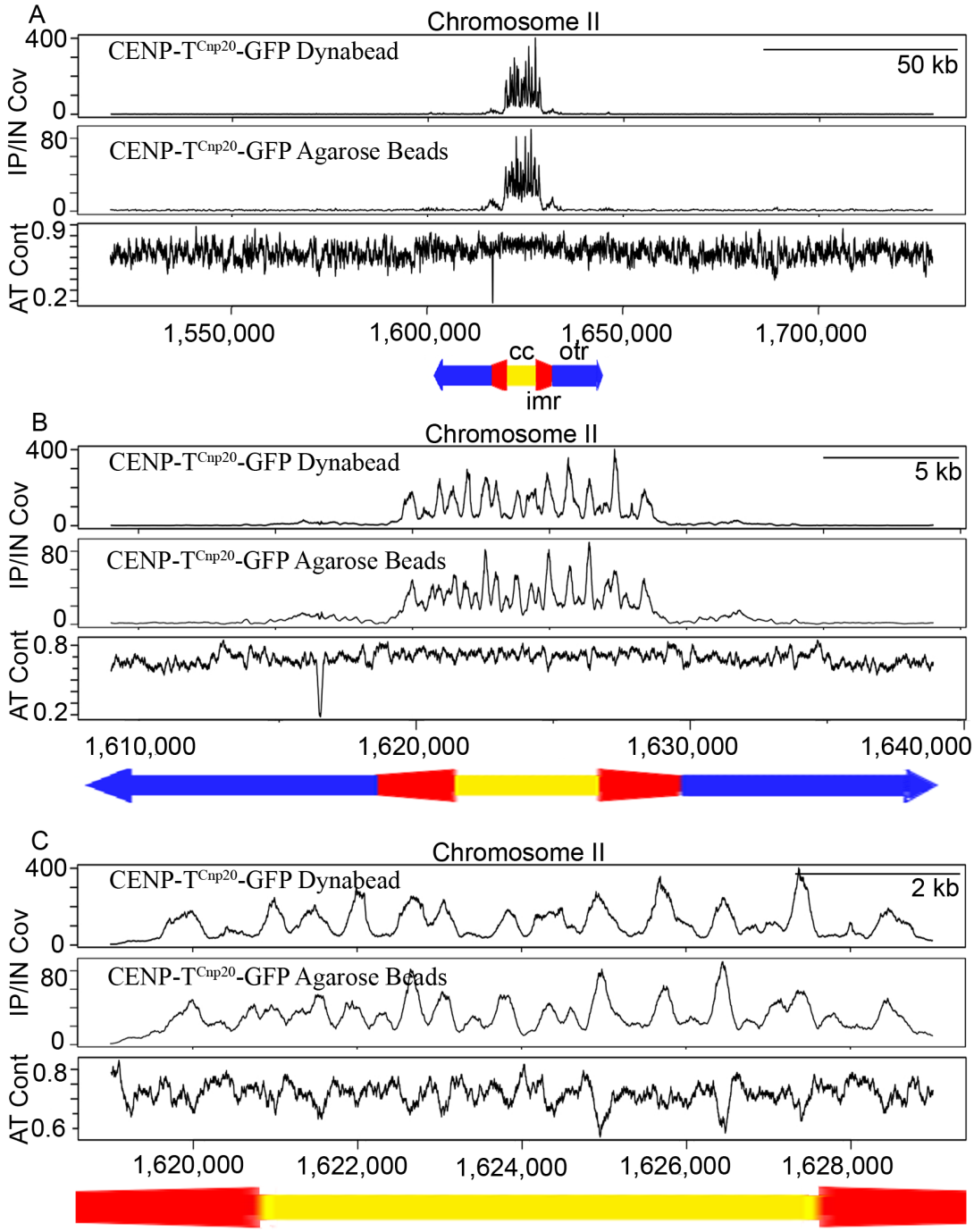


Figure 3.15. Agarose bead and Dynabead ChIP-seq result in slightly different enrichment patterns. CENP-T^{Cnp20}-GFP ChIP-seq performed using Dynabeads or agarose beads are shown with IP/IN coverage plotted across centromere II, including (A) 100 kb, (B) 10 kb, and (C) 0 kb around the central domain. ChIP-seq using agarose beads identifies the same peaks of enrichment present in the Dynabead data, but also shows additional enrichment peaks not observed before within the previously identified gaps between enrichment peaks.

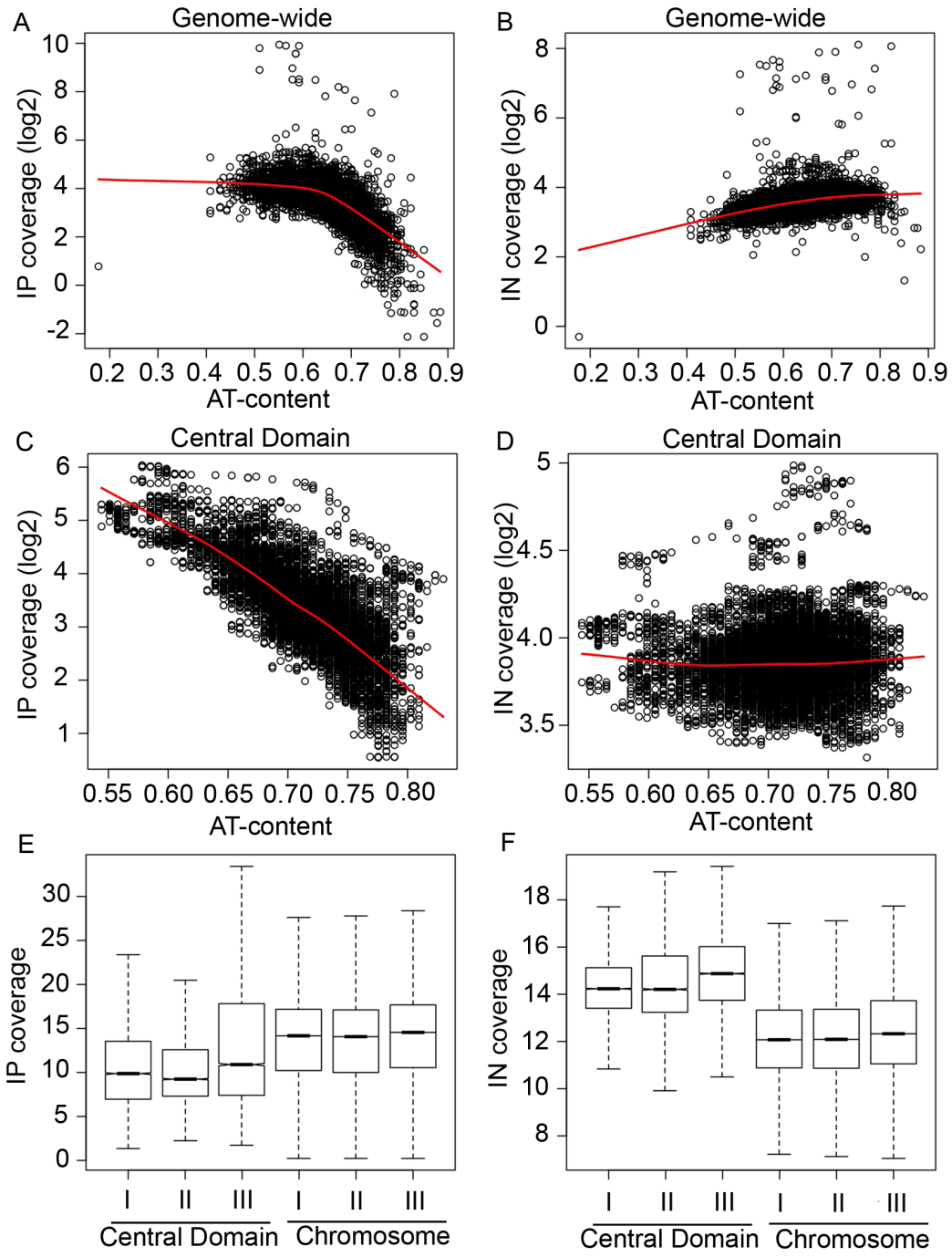


Figure 3.16. Bias in agarose bead ChIP-seq. Following illumina sequencing of a mock IP using agarose beads there is an observable lower coverage of regions with higher AT-content across the genome (A) following the IP (τ=-0.523), but (B) this bias is not observed in the IN data (τ=0.365). Within the central domain there is a similar negative correlation between AT-content and coverage of the IP (τ =-0.596), but not the IN data (τ=0.009). Lower coverage of the centromere central domain relative to the chromsomal average is observed in the (E) IP, but not in the (F) IN.

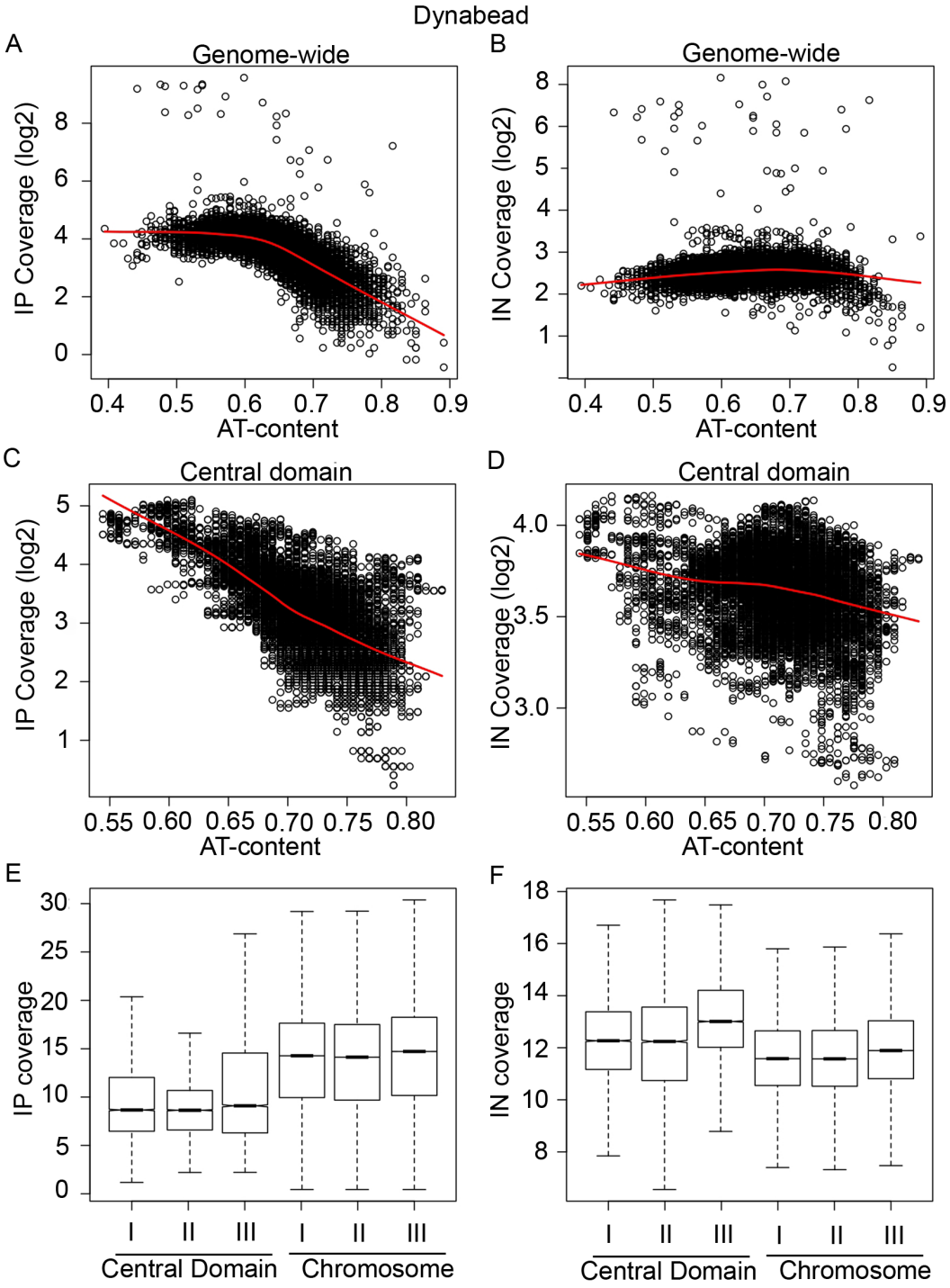


Figure 3.17. Bias in Dynabead ChIP-seq. Following illumina sequencing of a mock IP using magnetic Dynabeads and whole cell extract IN there is an observable lower coverage of regions with higher AT-content across the genome (A) following the IP (τ=-0.560), but (B) this bias is not observed in the IN (τ =0.101). (C) Similarly, within the central domain there is a negative correlation between AT-content and IP coverage (τ=-0.469), but (D) this bias is not observed in the IN coverage (τ=-0.159). Lower coverage of the centromere central domain relative to the chromosomal average is observed in the (E) IP, but not in the (F) IN.

domain can be detected using this methodology. Both methodologies have similar levels of bias so it remains unresolved as to why a second class of CENP-T^{Cnp20} peaks are present in agarose bead but not Dynabead ChIP-seq data.

It would be ideal to perform CENP-A^{Cnp1} ChIP-seq using Dynabeads to compare since that protocol results in higher enrichment with lower background. GFP-CENP-A^{Cnp1} cannot be used since this results in disrupted nucleosome positioning or additional crosslinking due to the large GFP tag. ChIP-seq using strains expressing HA-tagged CENP-A^{Cnp1} using magnetic Dynabeads could be undertaken to clarify the discrepancy. The primary highly positioned peaks are the high confidence occupancy pattern for CENP-A^{Cnp1} nucleosomes given their high enrichment in all ChIP-seq data sets, but additional positioned CENP-A^{Cnp1} nucleosomes may be present at secondary sites within the defined gaps. Regardless, both the primary and secondary sites still do not occupy all sites that are predicted by the nucleosome occupancy algorithm.

3.2.7. *S. cryophilus, S. octosporus,* and *S. pombe* have similar overall centromere organisations

S. pombe CENP-A^{Cnp1} nucleosomes are highly positioned in a sequence dependent manner within centromeres with large inter-nucleosomal gaps. Thus far it is unclear whether this reflects a conserved feature of centromeres that may be important for centromere function. Two related fission yeast species were analysed to explore this possibility. Centromeres in the fission yeast species *Schizosaccharomyces octosporus* and *Schizosaccharomyces cryophilus* are conserved in position but share no detectable sequence homology and thus likely represent evolutionarily ancient centromeres that lack conserved sequences (Rhind et al., 2011).

CENP-A^{Cnp1} and H3K9me2 ChIP-seq had previously been performed on chromatin extracted from *S. octosporus* and *S. cryophilus* using CENP-A^{Cnp1} anti-serum and agarose beads (H. Berger; Allshire lab unpublished), but was repeated to minimize sequencing bias that could potentially affect the analyses. All sequencing data was high quality and analysed as above. The centromeres had not been fully assembled in the genome assemblies of these species and *de novo* assembly from the CENP-A^{Cnp1} ChIP-seq data was not successful (data not shown).

Mapping of these ChIP-seq data to the assembled genomes identified areas of enrichment consistent with the presence of partial centromere assemblies on supercontig 5.3 in *S. octosporus* and supercontig 3.4 in *S. cryophilus*. In both species the centromere region is adjacent to the predicted centromere location and has conserved synteny of flanking genes. In both species the ChIP-seq data reveals the presence of non-overlapping regional CENP-A^{Cnp1} and H3K9me2 domains that are similar in arrangement to the *S. pombe* domains of CENP-A^{Cnp1} with flanking heterochromatin (Figure 3.18 and 3.19).

The genome-wide AT-content of *S. octosporus* and *S. cryophilus* are 62.5% and 62.3%, respectively. *S. octosporus* centromeres have an AT-content of 65.8% and the mapped CENP-A^{Cnp1} domains have a similar AT-content of 65.7%. The *S. cryophilus* centromere has an AT-content of 66.9% and AT-content of the mapped CENP-A^{Cnp1} domains is 67.7%. Similar to *S. pombe*, centromeres in both of these species are more AT-rich than the genome-wide average.

3.2.8 CENP-A^{Cnp1} positioning is similar in *S. octosporus* and *S. pombe*

Next it was determined whether CENP-A^{Cnp1} positioning sequence features were also shared between *S. octosporus, S. cryophilus,* and *S. pombe*. The high resolution positioning of CENP-A^{Cnp1} was determined in these species relative to the underlying sequence features.

In *S. cryophilus* there is high and relatively uniform enrichment of CENP-A^{Cnp1} across the assembled centromere region with no distinct peaks and gaps, negating further analysis (Figure 3.18). This may be due to a lower quality assembly of this region, as indicated by multiple short gaps in the assembly. In contrast, *S. octosporus* centromeres displayed highly positioned CENP-A^{Cnp1} peaks with intervening gaps (Figure 3.20). The new *S. octosporus* CENP-A^{Cnp1} ChIP-seq data showed less distinct peaks with higher enrichment even between peaks, similar to what was observed for *S. pombe*. However, large gaps were present between CENP-A^{Cnp1} peaks (median=363 bp). Furthermore, the CENP-A^{Cnp1} peaks occupy less AT-rich sequences (61.6%), while the gaps between nucleosomes are more AT-rich (67.1%; $p<0.001$). The peaks are highly positioned on sequences with greater CG-richness and AT-richness increases with increasing distance from these peaks. Additionally, motifs present within the S. octosporus CENP-A^{Cnp1} peaks and

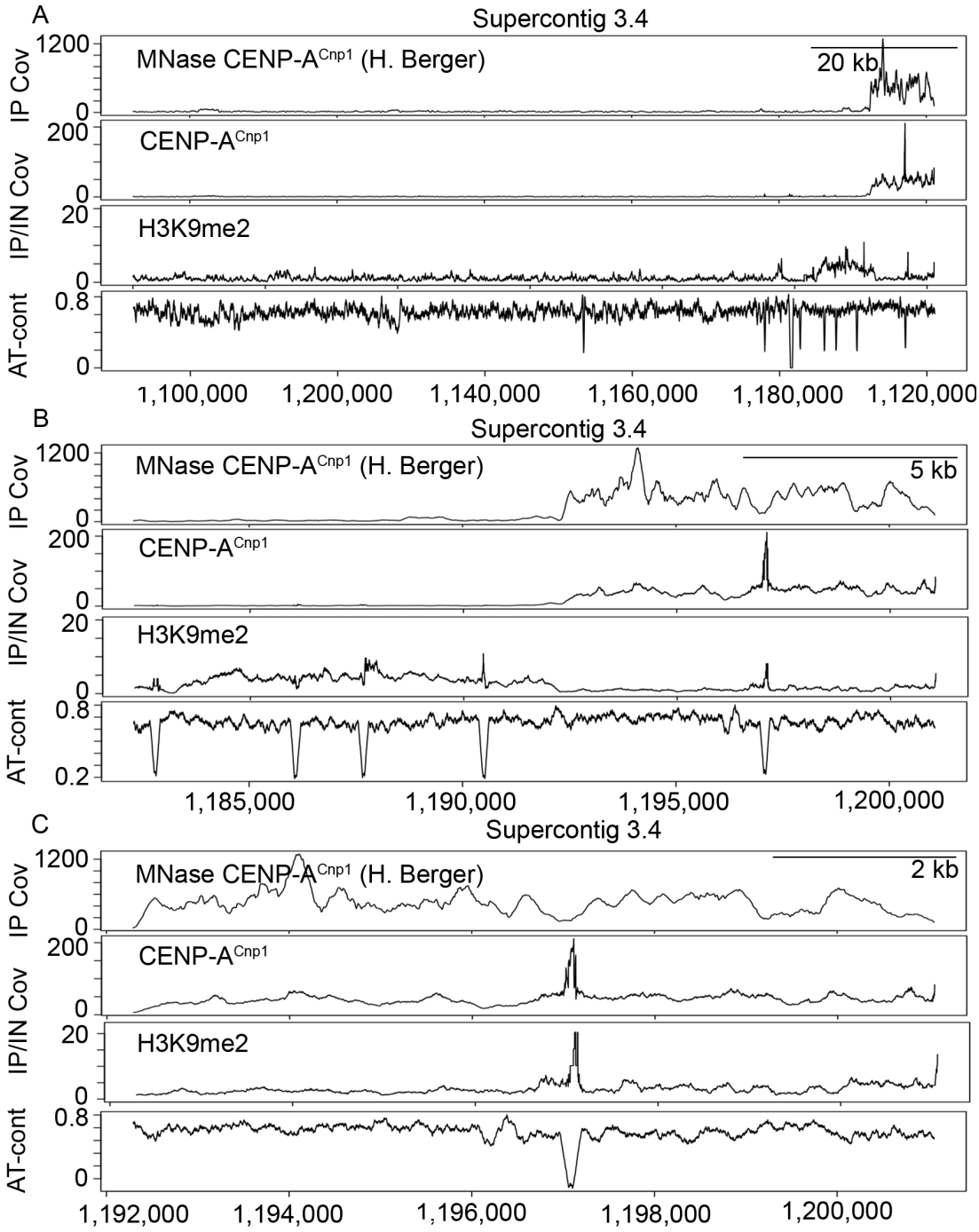


Figure 3.18. *S. cryophilus* centromere organisation. The candidate centromere on supercontig 3.4 of *S. cryophilus* is visualised with (A) 100 kb, (B) 10 kb, and (C) 0 kb of flanking sequences around the CENP-A^{Cnp1} domain. CENP-A^{Cnp1} ChIP-seq data using sonication sheared chromatin and MNase digested chromatin (H. Berger) both showed high enrichment at a regional domain. This region was flanked by a domain of enrichment for the heterochromatin associated mark H3K9me2. Regions of extremely low AT-content contain gaps with unknown sequence content.

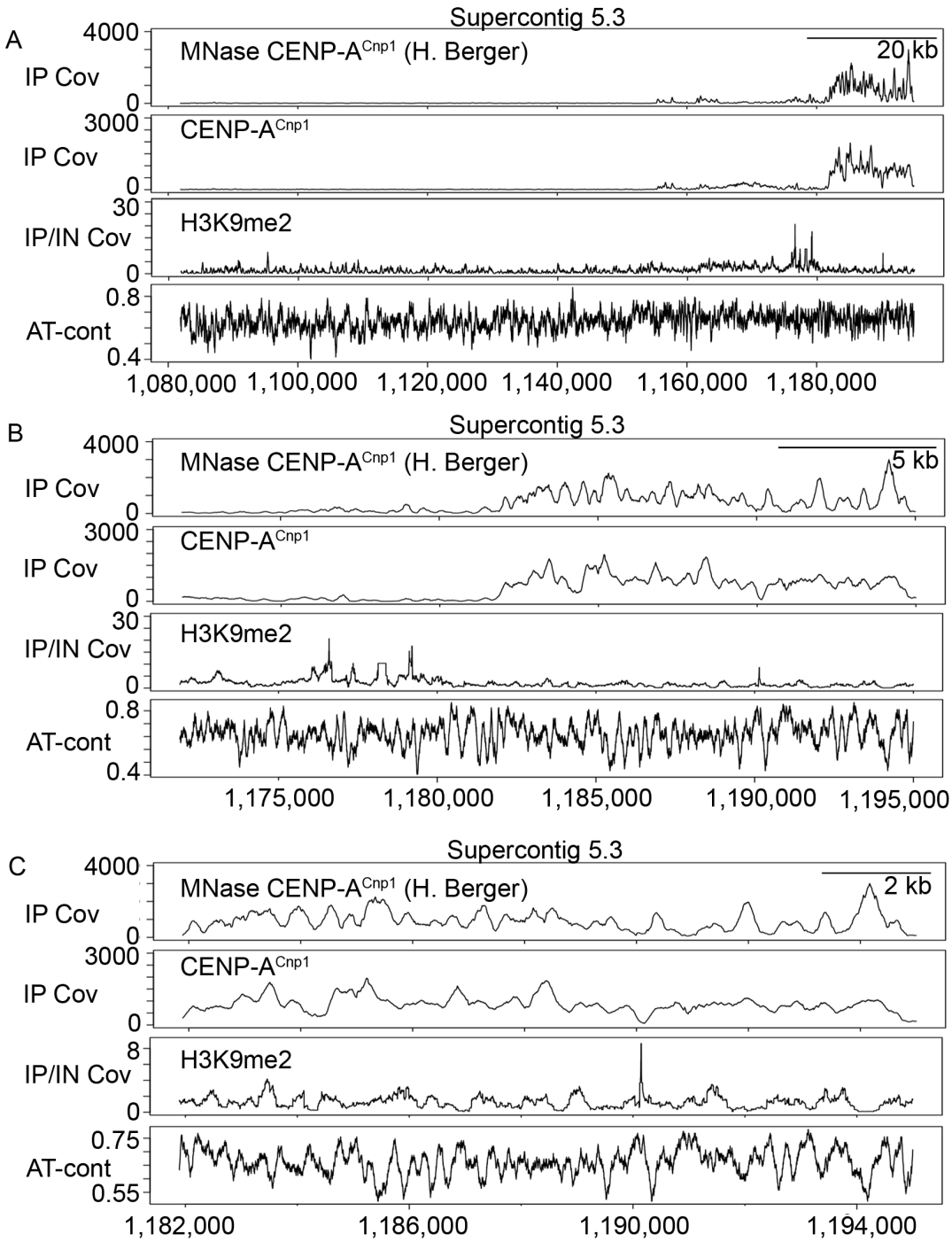


Figure 3.19. *S. octosporus* centromere organisation. The candidate centromere on supercontig 5.3 of *S. octosporus* is visualised with (A) 100 kb, (B) 10 kb, and (C) 0 kb of flanking sequences around the CENP-A^{Cnp1} domain. CENP-A^{Cnp1} ChIP-seq data using MNase digested chromatin (H. Berger) and sonication sheared chromatin both showed high enrichment at a regional domain. This region was flanked by a domain of enrichment for the heterochromatin associated mark H3K9me2.

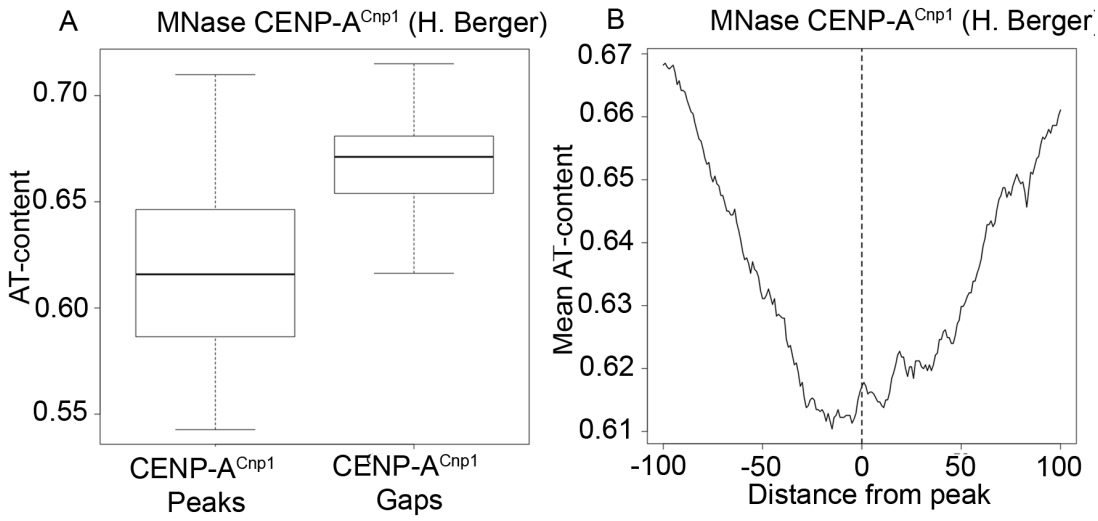


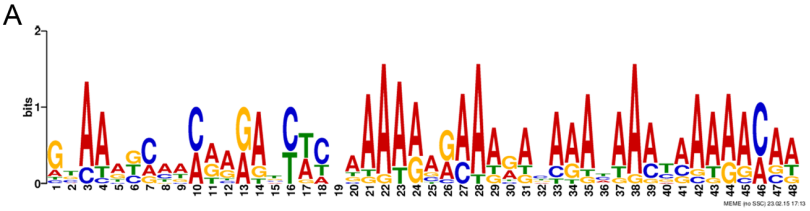
Figure 3.20. *S. octosporus* CENP-A is positioned by sequence features. (A) CENP-A^{Cnp1} occupies sites that have a lower AT-content (61.6%) than the more AT-rich gaps that separate CENP-A^{Cnp1} peaks (67.1%; $p<0.001$). (B) CENP-A^{Cnp1} peaks are positioned at a local minima of AT-content with AT-content increasing with increased distance from the CENP-A^{Cnp1} peak.

gaps were identified using MEME (Bailey and Elkan, 1994). The gaps between CENP-A^{Cnp1} nucleosomes were enriched for a 48 bp asymmetrical A/T rich motif, which is also purine rich, found in 12 of the 22 gaps with an E-value of 3.2e-1 (Figure 3.21). The CENP-A^{Cnp1} peaks were enriched for a 49 bp motif found in only 6 of the 23 peaks with an E-value of 9.3e-8 with unclear relevance (Figure 3.22). Flanking asymmetric AT-rich sequences lead to positioning of CENP-A^{Cnp1} in both *S. octosporus* and *S. pombe*. This may indicate that these sequence features might be important for centromere identity or function. However, *S. cryophilus* has a different pattern of CENP-A^{Cnp1} occupancy. This may result from a partial and potentially inaccurate assembly of the *S. cryophilus* centromere. Future de novo assembly of all centromeres in these species using PacBio sequencing will provide more insight into this question.

3.2.9. CENP-A^{Cnp1} occupancy profiles at neocentromeres differ from canonical centromeres

The above analysis shows that sequence features that correlate with highly positioned CENP-A^{Cnp1} nucleosomes are present at both *S. pombe* and *S. octosporus* centromeres despite the fact that the centromere DNA sequences in the species are not homologous. In addition to the conservation of these features, it is also of interest whether these features are strictly necessary for centromere function. To address this question CENP-A^{Cnp1} occupancy and positioning in strains carrying neocentromeres were analysed. Deletion of the canonical chromosome I centromere leads to the formation of a neocentromere near the left (strain cd39) or right telomere (strain cd60) of chromosome I on sequences with no shared homology to the endogenous centromeres (Ishii et al., 2008; Ogiyama et al., 2013). Prior to neocentromere formation the sequences on which neocentromeres form are not under the same selective pressure as endogenous centromere sequences and are thus unlikely to possess all features or optimal centromere activity. CENP-A^{Cnp1} ChIP-seq was previously performed (unpublished, H. Berger; Allshire lab) in strains carrying either the neocentromere cd39 (left telomere) or cd60 (right telomere). CENP-A^{Cnp1} ChIP-seq was newly performed in a strain carrying neocentromere cd39 to verify the data. Since these neocentromeres are novel they have not been under selective pressure towards optimal centromere sequence organization. Analysis will provide further information on the features that are necessary for centromere function.

A

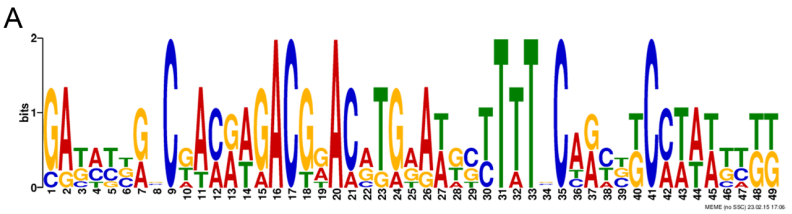


B

Name	Strand	Start	*p*-value	Sites
16	+	568	7.71e-22	AGAGTCCTTG GTAAAGCAACGAGATCTCAAAAAGAGAAAGATAAATAAATAAAAACAA TTAAAAAAAG
15	+	545	9.32e-21	AGAGTCCTTG GTAAAGCAACGAGATCTCAAAAAGAGAAAGATAAATAAATAAAAACAT TTTCTTTCTA
10	-	52	2.78e-12	GATTATGAAA AGAATGCTACTAAATCTTCGAAAAGAAAGGACAAAAAAATCAAAAACA GAAAGCGACG
9	-	116	8.39e-11	AGCAGAATAT GTAAATGAGCAAGAGTAAAAAGAGCCAATGAAAAAAAAAAAAAACCAA ACAAATGAAA
21	+	47	9.81e-11	GCATAATCTG AGAACCCCTCTTGGTTATGAAAAAAGCAATGTAAACAAACAAAGAAAA GCCTGAAGCG
13	+	153	1.24e-10	CCATACCACA CCATGGCCTAAAGTGTTCTATAAAAGAAAAACTGAAAACTGAAAACAG AGACAGAGAC
14	-	642	1.67e-10	GTAAAGTTTT GACAGTGCGAGGGAACTCCTAATAGAAAAAAGAAGCAGATAATAACAA ATATGAAATA
5	+	49	2.25e-10	TCCCGTCGTC GCCTACCATAGAAATCGCGCAAAATAAAAAAGAAACGAAGCAAGGAGG TATGACCGTA
3	-	97	3.23e-10	TTGGTTACTG GGAAGGAGCAAGGAGCAAGGAAAAAGCATAAAAGATTAGCAATAGCGA TCAGCACGTA
6	+	82	4.59e-10	GTATATAATT TTAAATAAACACAAATTCCAAATAAAAAAGGCCAAGTACCAAGAACCA TATTACGGGG
20	-	591	7.89e-10	TATATCAAAC GTACACCTACAAAGCCATATAAAACGATAGTCCGTCAACGCGAGACAT ATGTAATATT
4	+	255	2.05e-09	GTACGATTGA GCAATTAAAAAGAACTACTAGAAGAGCAGCTCATATAAATGCAAAAAT GTTAGATATT

Figure 3.21. Gaps between CENP-A^{Cnp1} in *S. octosporus* are asymmetric A/T rich. The program MEME (Bailey and Elkan 1994) was used to identify motifs enriched within the gaps between CENP-A^{Cnp1} nucleosomes in *S. octosporus*. (A) The most significant motif identified was a 49 bp asymmetric A/T rich motif with an E-value of 3.2e-001. (B) This motif was found at 12 of the 22 CENP-A^{Cnp1} peaks.

A



B

Name	Strand	Start	*p*-value	Sites
16	+	47	4.08e-21	AACAACTCAA GATACGAGCGACGTGACGGACGTGAATTGCTTTGCAACGTCCTTTTTGG CTGATTTGTT
15	+	47	4.08e-21	AATAACTCAA GATACGAGCGACGTGACGGACGTGAATTGCTTTGCAACGTCCTTTTTGG CTGATTTGTT
22	-	59	2.12e-19	GCACTTGGTT GAGCTCGCCTAAAAGACGAACATGGAAGCTTTTCCAGTTGCAAAAGCTT AACGATAGAG
1	+	67	2.12e-19	GCACTTGGTT GAGCTCGCCTAAAAGACGAACATGGAAGCTTTTCCAGTTGCAAAAGCTT AACGATAGAG
12	+	84	7.74e-15	AATGTAGTTC CGCAGTGTCGTCGAAACGTACAGAAGTGCTTTTACTGCCTCCTAATTTT GAATTTTCTA
6	+	71	2.11e-14	TGCTGATAGC GATTTTGACAACAAGACTGAACTGTAAATTTATTCCAACGCCTATCAGG AATTAATTGT

Figure 3.22. CENP-A^{Cnp1} enrichment peaks have a shared motif with unclear relevance in *S. octosporus*. The program MEME (Bailey and Elkan 1994) was used to identify motifs enriched at CENP-A^{Cnp1} peaks in S. octosporus. (A) The most significant motif identified was a 49 bp motif with an E-value of 9.3e-008. (B) This motif was found only at 6 of the 23 CENP-A^{Cnp1} peaks. This analysis is limited by the small number of peaks available for analysis.

Previously collected CENP-A^{Cnp1} ChIP-seq data shows high enrichment proximal to the left and right telomeres in strains cd39 (Figure 3.23) and cd60 (Figure 3.24), respectively. CENP-A^{Cnp1} ChIP-seq was repeated on the cd39 strain and shows similar enrichment, but has a more uniform high level of enrichment across the entire CENP-A^{Cnp1} domain with less well-defined peaks. This is similar to what was observed with the new CENP-A^{Cnp1} ChIP-seq data at endogenous centromeres. However, at the cd39 neocentromere enrichment of CENP-A^{Cnp1} occurs over a larger domain than that observed at any of the thee endogenous centromeres, with approximate CENP-A^{Cnp1} domain sizes of 13.7 kb and 22.6 kb at neocentromeres cd39 and cd60, respectively. This is in agreement with previous work (Ishii et al., 2008). In particular, neocentromere cd60 forms over a larger domain and contains several CENP-A^{Cnp1} enriched regions with large regions of low enrichment between. As ChIP is performed on a population of cells, it is possible that the larger domain represents either the oscillation of the neocentromere between multiple sites in different cells or the formation of a single larger chromatin domain due to suboptimal sequence features.

3.2.10. Neocentromeres do not have sequence features that promote highly positioned nucleosomes

Based on the previously collected MNase digested CENP-A^{Cnp1} ChIP-seq data neocentromere cd39 CENP-A^{Cnp1} peaks are less AT-rich (64.2% AT-content) than the gaps between CENP-A^{Cnp1} peaks (69.6% AT-content; $p<0.001$; Figure 3.25). Using the current peak calling method it was not possible to define peaks for the new cd39 CENP-A^{Cnp1} ChIP-seq data due to the high enrichment across the entire CENP-A^{Cnp1} domain, so this data was not used for further analyses. At neocentromere cd60 CENP-A^{Cnp1} occupies less AT-rich sequences (62.4% AT-content) than the gaps between peaks (67.2% AT-content; $p<0.001$). This pattern is similar to that observed at endogenous centromeres, although the neocentromeres are overall less AT-rich. The median gap size between CENP-A^{Cnp1} nucleosomes is 387.5 bp for neocentromeres cd39 and 433 bp for neocentromere cd60. MEME was used to identify motifs present in the CENP-A^{Cnp1} peaks and the gaps between peaks (Bailey and Elkan, 1994). However, no significantly enriched motifs were present. This is in striking contrast to asymmetric AT-rich motifs identified in the gaps between CENP-A^{Cnp1} nucleosomes at both *S. pombe* and *S. octosporus*

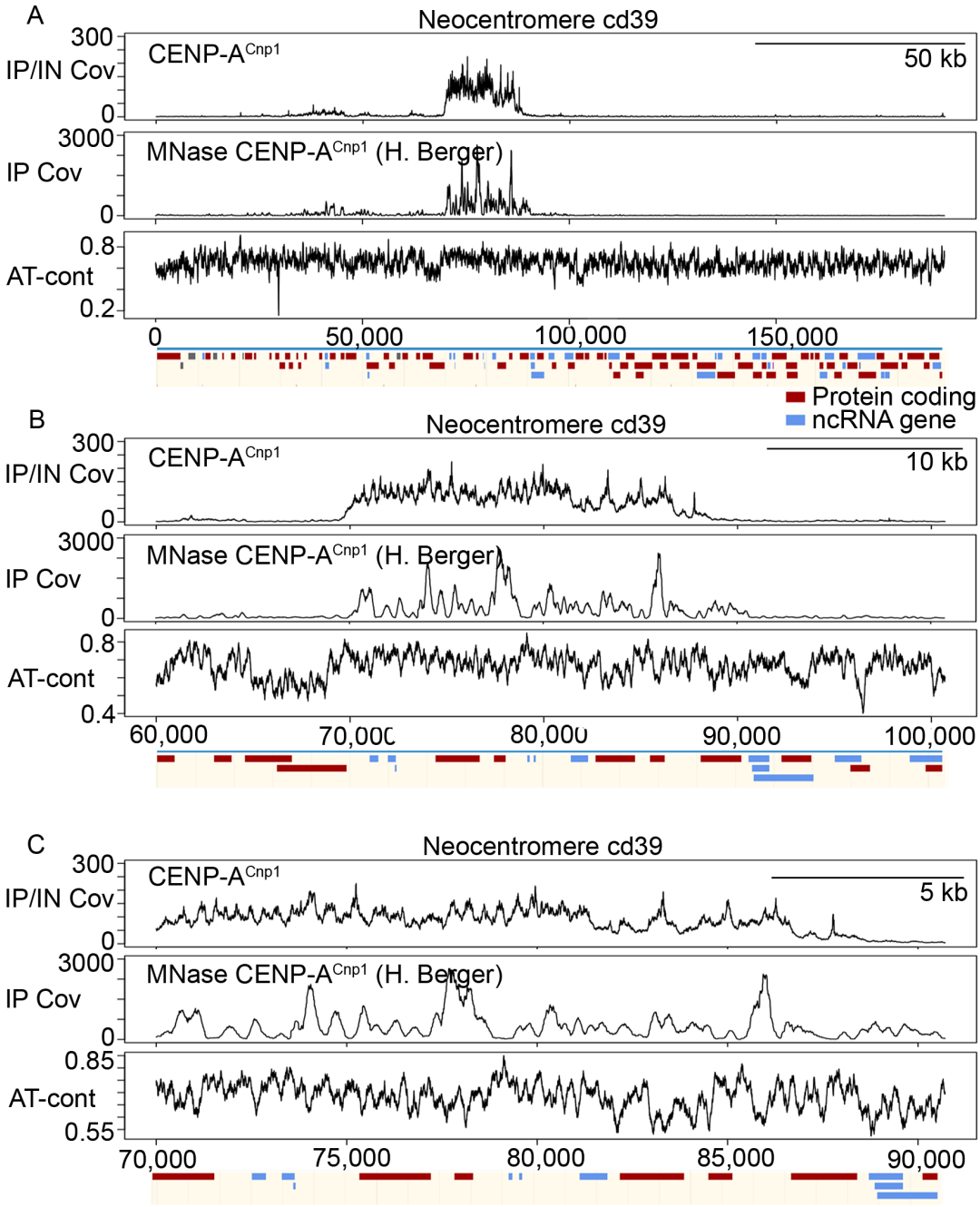


Figure 3.23. Centromere organisation at neocentromere cd39. *S. pombe* neocentromere cd39 on chromosome I is visualised with (A) 100 kb, (B) 10 kb, and (C) 0 kb of flanking sequences around the CENP-A^{Cnp1} domain. CENP-A^{Cnp1} ChIP-seq data using MNase digested chromatin (H. Berger) and sonication sheared chromatin both showed high enrichment of CENP-A^{Cnp1} at a regional domain.

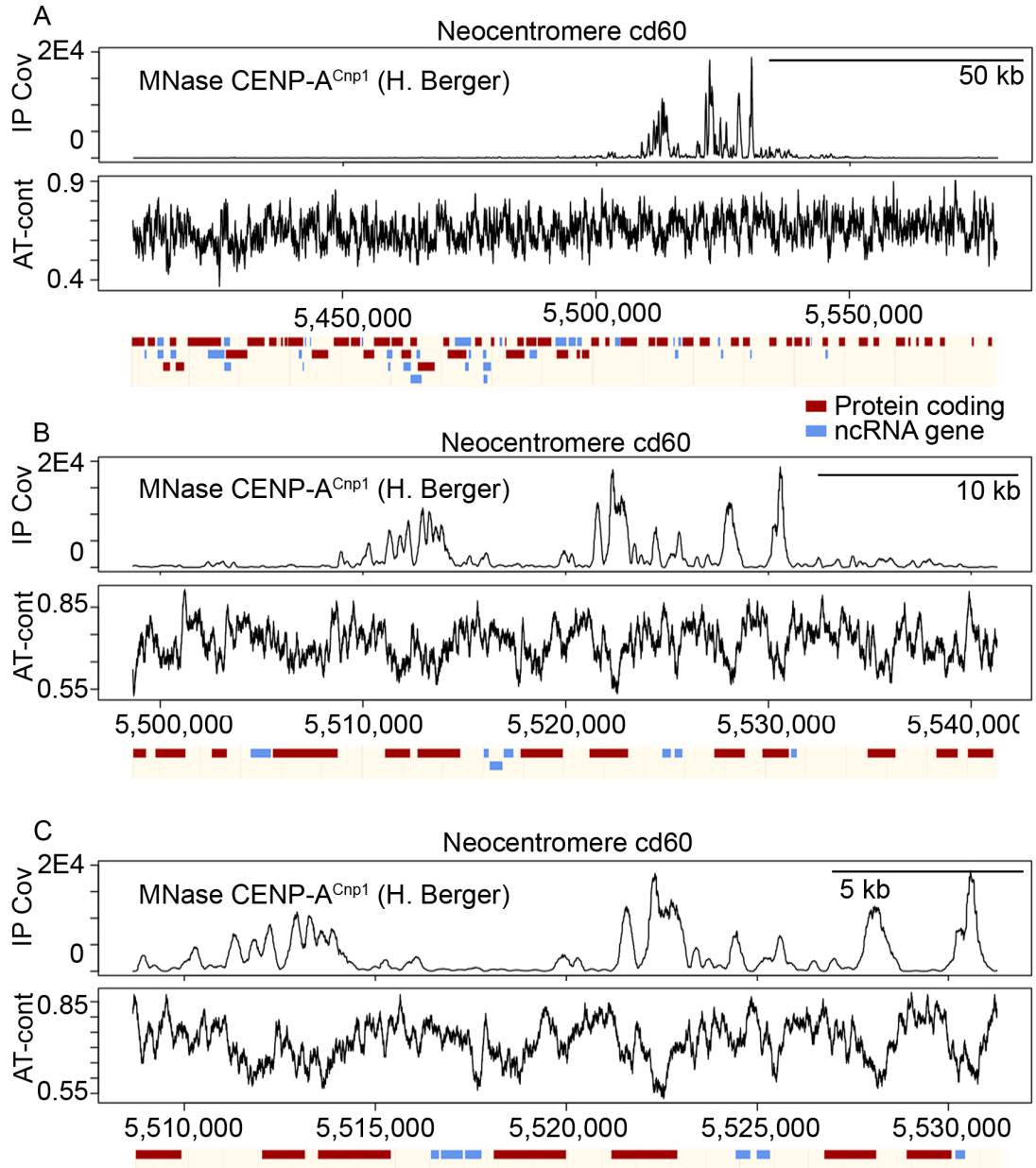


Figure 3.24. Centromere organisation at neocentromere cd60. *S. pombe* neocentromere cd60 on chromosome I is visualised with (A) 100 kb, (B) 10 kb, and (C) 0 kb of flanking sequences around the CENP-A^{Cnp1} domain. CENP-A^{Cnp1} ChIP-seq data using MNase digested chromatin (H. Berger) shows high enrichment at a regional domain composed of multiple regions of CENP-A^{Cnp1} enrichment separated by large regions without enrichment.

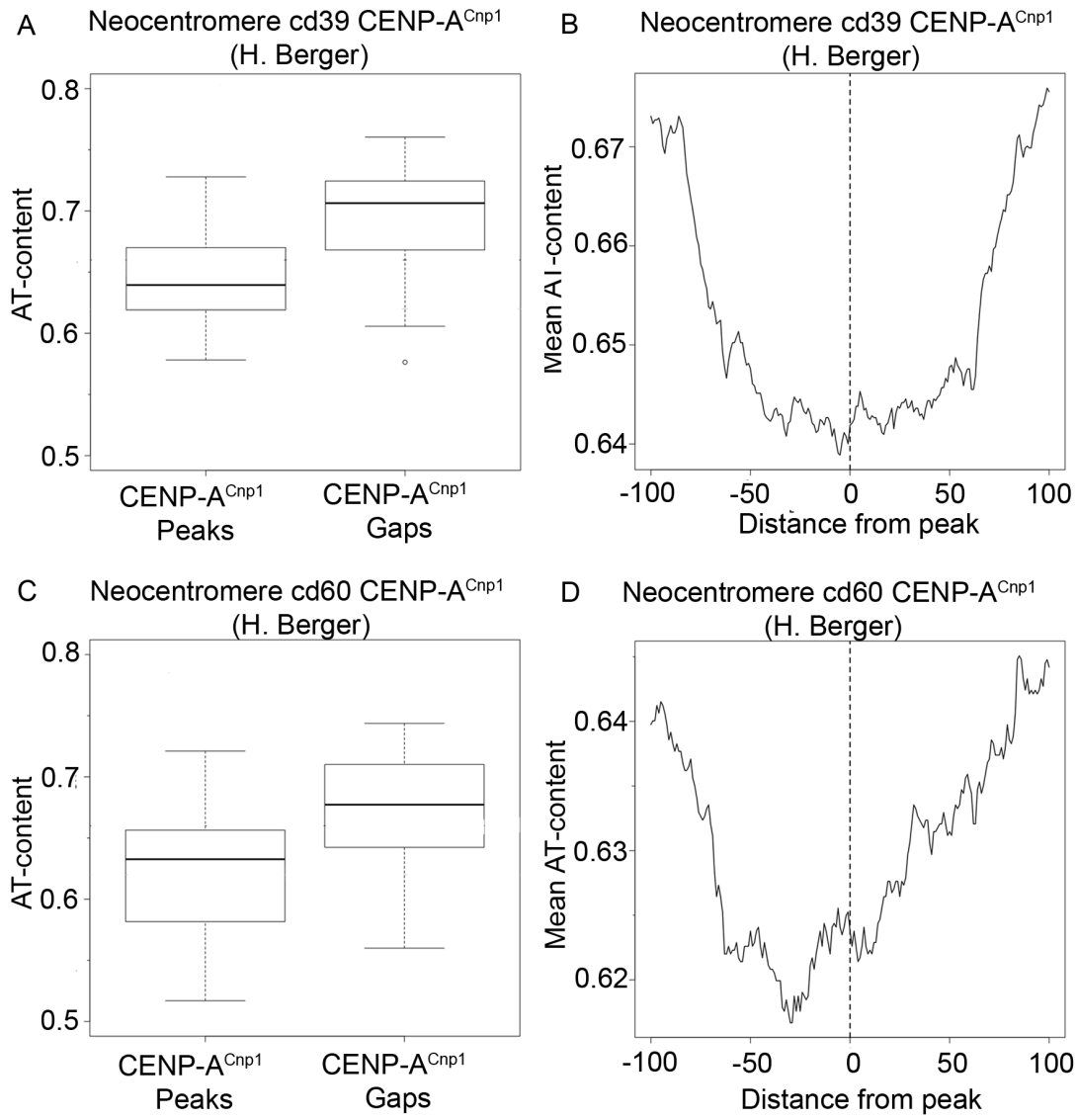


Figure 3.25. CENP-A^{Cnp1} occupancy differs between neocentromeres and canonical centromeres. (A) Neocentromere cd39 CENP-A^{Cnp1} peaks occupy sites that have lower AT-content (64.2%) than the gaps between peaks (69.6%; $p<0.001$). (B) Unlike at canonical centromeres, AT-content does not increase with increasing distance from the peak, but rather remains relatively constant. (C) Neocentromere cd60 CENP-A^{Cnp1} peaks occupy sites that have lower AT-content (62.4%) than the gaps between peaks (67.2%; $p<0.001$). (D) Unlike at canonical centromeres, there is no clear pattern of increasing AT-content as distance from the peak increases.

endogenous centromeres. An analysis of the distribution of k-mers is being undertaken to investigate the sequence features at neocentromere in greater detail.

The gaps between CENP-A^{Cnp1} nucleosomes at neocentromeres are more AT-rich than the CENP-A^{Cnp1} occupied sites, but the pattern of sequence features around CENP-A^{Cnp1} peaks differs from endogenous centromeres (Figure 3.25). The AT-content around CENP-A^{Cnp1} occupied sites at neocentromere cd39 is different from the endogenous centromeres. At neocentromere cd39 CENP-A^{Cnp1} peaks occupy sites of lower AT-content, but the sites surrounding the peaks also have a similarly low AT-content. This suggests that CENP-A^{Cnp1} nucleosomes at these sites are not strongly positioned by the sequence features and that adjacent sites could also be occupied with a similar preference. The CENP-A^{Cnp1} peaks detected at neocentromere cd60 also did not occur at local minimum of AT-content.

The factors influencing CENP-A^{Cnp1} positioning at neocentromeres are similar to that at endogenous centromeres. At both sites CENP-A^{Cnp1} peaks over the less AT-rich sequences. However, there is also some indication that neocentromeres contain suboptimal sequence features for centromere formation. At canonical centromeres CENP-A^{Cnp1} is highly positioned by flanking AT-rich sequences, while at least one neocentromere this is not the case.

3.3 Discussion

The results presented above indicate that CENP-A^{Cnp1} occupancy is highly correlated with underlying DNA sequence features and that sequence could be, but is not necessarily, primarily responsible for positioning CENP-A^{Cnp1} within the central domain regions of centromeres. This is defined primarily by a preference for occupying sites that are less highly AT-rich. This preference is shared with canonical H3-containing nucleosomes and both CENP-A^{Cnp1} and H3 nucleosomes are predicted to occupy the same sites within the centromeres. However, not all predicted sites are occupied by CENP-A^{Cnp1} *in vivo*, so other mechanisms and processes also affect CENP-A^{Cnp1} occupancy. Despite the overall high AT-content of the *S. pombe* centromeres, the mechanism that directs CENP-A^{Cnp1} to centromeres for incorporation is unlikely to rely on such specific sequence features, since the features that affect CENP-A^{Cnp1} occupancy are also shared with canonical H3 nucleosomes. In addition, the specific enrichment of CENP-A^{Cnp1} and depletion of

H3 at centromeres are unlikely to be attributed to sequence features alone since they have similar sequence preferences. However, it remains possible that, although both prefer similar sequences, some subtle difference in the strength of this preference results in differential turnover of H3 relative to CENP-A^{Cnp1} particles at centromeres.

The overall organisation of CENP-A^{Cnp1} nucleosomes at centromeres is also conserved with a related species despite the absence of similarity. The genome of the fission yeast *S. octosporus* contains a partially assembled AT-rich centromere. A regional CENP-A^{Cnp1} domain with flanking heterochromatin is evident. Similar to *S. pombe*, *S. octosporus* CENP-A^{Cnp1} nucleosomes are highly positioned on the more CG-rich sequences with more AT-rich sequences residing between nucleosomes. Thus, these sequence features may represent a generally conserved aspect of centromere sequence organisation in fission yeasts. Discretely positioned nucleosomes separated by large gaps may be beneficial for kinetochore attachments and centromere function. These analyses could only be applied to a partial assembly of one centromere from *S. octosporus,* but sequencing by PacBio is currently being undertaken to assemble all three centromeres in both *S. octosporus* and *S. cryophilus* (P. Tong, Allshire lab, personal communication).

However, the above sequence features are not strictly required for kinetochore formation and centromere function. Formation of evolutionarily young neocentromeres takes place at loci adjacent to the telomeres of chromosome I (Ishii et al., 2008; Ogiyama et al., 2013). At these neocentromeres CENP-A^{Cnp1} occupies domains that are larger than the CENP-A^{Cnp1} domains at canonical centromeres. These neocentromeres are relatively AT-rich, though not to the same extent as endogenous centromeres, and the pattern of AT-content differs across neocentromeres. The evolutionarily young neocentromeres lack the variable regions of AT-content that lead to discrete highly positioned CENP-A^{Cnp1} nucleosomes. Since neocentromeres are evolutionarily young it is possible that they represent suboptimal sequences for centromere formation and may evolve towards acquiring the characteristics of canonical centromeres. This possibility could be tested with a competition assay using progressively mutagenized minichromosomal neocentromeres to determine if increases in fitness are associated with more strongly positioned CENP-A^{Cnp1} nucleosomes. An alternative test would be to

investigate the evolution of a naturally occurring ancient neocentromere if a suitable wild-strain with such a centromere was discovered.

The highly positioned CENP-A^{Cnp1} nucleosomes in fission yeast are not unique. At the point centromeres of the budding yeast (Saccharomyces cerevisiae) CENP-A^{Cse4} nucleosomes are positioned in phase with flanking canonical nucleosomes (Bloom and Carbon, 1982; Yuan et al., 2005). Moreover, at human centromeres CENP-A nucleosomes are regularly phased as a result of interactions with CENP-B (Ando et al., 2002) and it is known that CENP-B dimers can bind to two CENP-B boxes to position a nucleosome on the intervening DNA loop *in vitro* (Yoda et al., 1998). It has been proposed that positioned CENP-A nucleosomes are an evolutionarily conserved feature of centromeres that may be important for proper kinetochore formation (Song et al., 2008). The specific selective force influencing this pattern remains unknown.

Chapter 4: Analysis of CENP-A^{Cnp1} distribution following CENP-A^{Cnp1} over-expression

4.1 Introduction

In most species it is important that a single centromere is formed on each chromosome. Multiple mechanisms contribute to the formation of a single functional centromere on each chromosome. CENP-A expression and incorporation are temporally regulated in a manner distinct from core histones. Core histones accumulate at high levels prior to entry into S-phase, during which incorporation into chromatin occurs following DNA synthesis (Harris et al., 1991; Marzluff et al., 2008). In contrast, fission yeast CENP-A^{Cnp1} accumulates earlier in the cell cycle during late M and G1/S (Takahashi et al., 2000) and is incorporated at the centromere during G2 (Lando et al., 2012). New CENP-A accumulates in human cells primarily during G2, meaning that CENP-A incorporation is temporally separate from DNA replication and CENP-A synthesis, and expression of CENP-A only during S-phase leads to a loss of centromere specific incorporation (Shelby et al., 1997, 2000). This temporal separation between core histone and CENP-A expression may allow for centromere specific targeting or more efficient CENP-A homotypic nucleosome formation (Shelby et al., 1997, 2000). In additional to the temporal control of CENP-A incorporation there is also distinct spatial targeting of CENP-A incorporation to achieve centromere specificity. Scm3 and the human ortholog HJURP act as chaperones for CENP-A and localise CENP-A specifically to centromeres (Barnhart et al., 2011; Foltz et al., 2009; Hayashi et al., 2004; Pidoux et al., 2009; Sanchez-Pulido et al., 2009; Shuaib et al., 2010).

Despite the presence of a dedicated chaperone and assembly mechanism for CENP-A there is plasticity in the chromosomal locations at which CENP-A is incorporated among epigenetically defined centromeres. Centromere loss from a chromosome can be resolved by formation of a neocentromere on non-centromeric sequences (Ishii et al., 2008; Marshall et al., 2008; Ogiyama et al., 2013; Shang et al., 2013). This balance between centromere specificity and neocentromere flexibility depends in part on levels of CENP-A in the cell. Over-expression of CENP-A leads to ectopic incorporation at non-centromeric loci (Castillo et al., 2013; Choi et al., 2012; Gascoigne et al., 2011; Van Hooser et al., 2001; Lacoste et al., 2014). In

budding yeast over-expression of CENP-A^{Cse4} leads to ectopic CENP-A^{Cse4} incorporation and subsequent segregation defects (Au et al., 2008). In human cells ectopic CENP-A incorporation leads to recruitment of a subset of kinetochore components but a functional kinetochore does not form (Gascoigne et al., 2011; Van Hooser et al., 2001; Lacoste et al., 2014). In fission yeast CENP-A ^{Cnp1} over-expression primarily leads to ectopic incorporation over the centromeric outer-repeats and subtelomeric regions, as well as a low level of incorporation at a subset of genes in FACT mutants (Castillo et al., 2013; Choi et al., 2012).

The ectopic incorporation of CENP-A ^{Cnp1} following over-expression in fission yeast was investigated previously using tiling arrays (Castillo et al., 2013; Choi et al., 2012). These experiments were constrained by the limited dynamic range and resolution inherent in tiling arrays. Therefore, I reinvestigated CENP-A ^{Cnp1} over-expression using ChIP-seq to examine ectopic CENP-A ^{Cnp1} incorporation at high resolution and sensitivity. This was used to identify features that might influence CENP-A ^{Cnp1} incorporation at ectopic sites and that might also contribute to CENP-A^{Cnp1} incorporation at canonical centromeres and the process of neocentromere formation.

4.2 Results

4.2.1 Over-expression of CENP-A^{Cnp1} leads to increased occupancy at the central domain and ectopic incorporation over outer repeats at centromeres

Analyses of CENP-A^{Cnp1} and H3 nucleosome occupancy show that they have similar sequence preferences (chapter 3). Within the central domain of centromeres CENP-A^{Cnp1} nucleosomes occupy essentially the same sites that canonical H3-containing nucleosomes are predicted to occupy. However, it is possible that that subtle difference in the strength of the sequence preferences may lead to the preferential incorporation of CENP-A^{Cnp1} at centromeres. CENP-A^{Cnp1} over-expression was used to test if particular sequence features genome-wide are associated with CENP-A^{Cnp1} incorporation, which could provide more information about its sequence preferences.

An *S. pombe* strain containing a plasmid integrated at the ars1 locus that allows moderate expression of CENP-A^{Cnp1} driven by an *nmt41* promoter was compared with a strain with wild-type CENP-A^{Cnp1} expression levels carrying an integrated

empty control plasmid containing only the *nmt41* promoter at ars1. The *nmt41* promoter is de-repressed by growth in the absence of thiamine (Forsburg, 1993). Growth for 48 hours in liquid minimal medium lacking thiamine induced expression of the integrated *nmt41* promoter. In the *nmt41*-CENP-A^{Cnp1} strain this has previously been observed to lead to high levels of CENP-A^{Cnp1} expression and ectopic incorporation at the outer repeats and subtelomeric regions (Figure 4.1; Choi et al., 2012). Levels of CENP-A^{Cnp1} could not be directly assessed since the available anti-CENP-A^{Cnp1} serum is unable to detect CENP-A^{Cnp1} by western analysis. GFP-CENP-A^{Cnp1} was previously used to quantify CENP-A^{Cnp1} levels (Choi et al., 2012), but this leads to compromised positioning or additional cross-linking and therefore is not ideal for use in high-resolution ChIP-seq mapping of CENP-A^{Cnp1} (Chapter 3). Instead over-expression of CENP-A^{Cnp1} was verified by the detection of a significant increase in the levels of CENP-A^{Cnp1} associated with the outer repeats at centromeres by ChIP-qPCR in the strain that should overexpress CENP-A^{Cnp1} compared to the wild-type expression control strain (Figure 4.1). Two biological replicates with two technical replicates were collected for the CENP-A^{Cnp1} over-expressing and control strains. CENP-A^{Cnp1} enrichment was detected within the central domain by ChIP-qPCR. CENP-A^{Cnp1} was enriched ~130 fold at one position tested in the central domain of the centromere on chromosome II relative to the actin gene (*act1*$^{+}$) in control cells, while CENP-A^{Cnp1} was enriched ~300 fold at the same position in cells over-expressing CENP-A^{Cnp1}. Cells over-expressing CENP-A^{Cnp1} have reduced extraction of central domain chromatin in the IN and therefore an increase in CENP-A^{Cnp1} occupancy within the central domain cannot be inferred from the change in CENP-A^{Cnp1} enrichment (discussed further below). However, this could potentially reflect an increase in CENP-A^{Cnp1} leading to higher occupancy of normally occupied sites or occupancy of the gaps between CENP-A^{Cnp1} where CENP-A^{Cnp1} normally is not enriched. CENP-A^{Cnp1} is not normally detected on the outer repeats at centromeres. Consistent with this, no enrichment was evident on the outer repeats in the control strain. However, CENP-A^{Cnp1} was enriched ~17 fold on the outer-repeats relative to *act1*$^{+}$ in cells over-expressing CENP-A^{Cnp1}.

These results are consistent with previous observations that CENP-A^{Cnp1} over-expression leads to ectopic incorporation at the outer repeats (Castillo et al., 2013). Even though the levels of CENP-A^{Cnp1} protein could not directly be examined, these

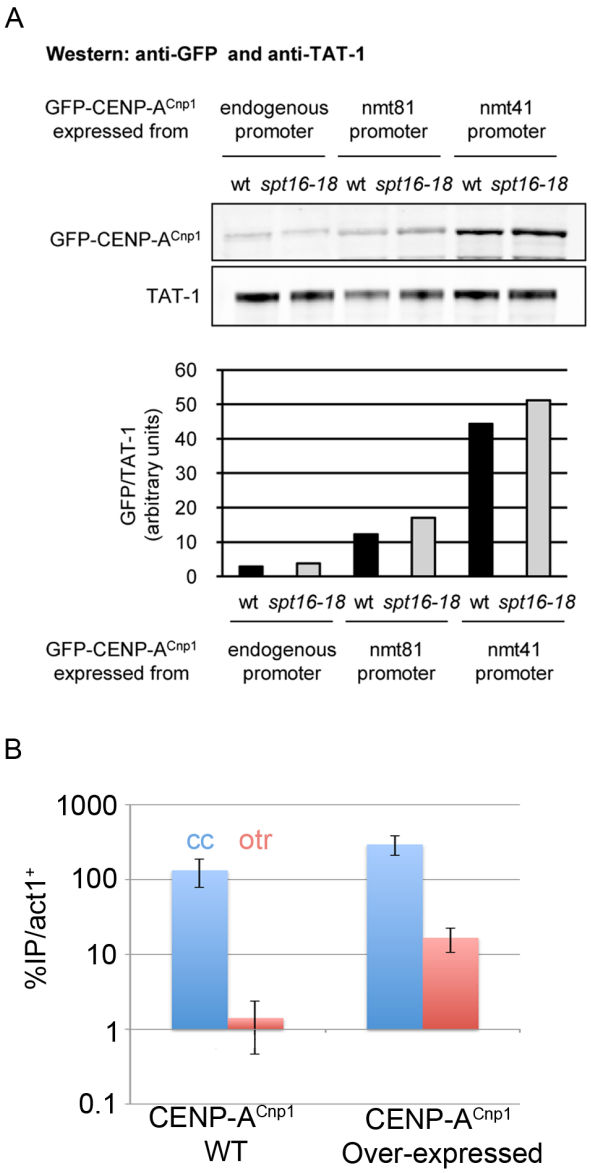


Figure 4.1. CENP-A^{Cnp1} over-expression leads to ectopic incorporation at the outer repeats. (A) From Choi et al. 2012, levels of CENP-A^{Cnp1} following over-expression were previously quantified using GFP-CENP-A^{Cnp1}. (B) Quantification by Western Analysis is not possible with wild-type CENP-A^{Cnp1}, which was necessary for this experiment, and therefore over-expression was verified by observing ectopic incorporation by qPCR. Within the central core (cc) CENP-A^{Cnp1} was highly enriched relative to the negative control actin (act1^{+}) locus in a strain with wild-type (WT) CENP-A^{Cnp1} expression and enrichment of CENP-A^{Cnp1} is higher in cells over-expressing CENP-A^{Cnp1}. At the outer-repeats (otr) there is no enrichment of CENP-A^{Cnp1} detected in wild-type cells, but there is enrichment (~23 fold) consistent with ectopic incorporation in cells over-expressing CENP-A^{Cnp1}.

initial analyses suggest that the CENP-A^{Cnp1} is over-expressed in these cells and that the samples can be used to look for ectopic incorporation genome-wide.

4.2.2 Ectopic CENP-A^{Cnp1} incorporation following overexpression is highly enriched at centromeres and subtelomeric regions

Ectopic incorporation of CENP-A^{Cnp1} following over-expression was investigated genome-wide by ChIP-seq of CENP-A^{Cnp1} using the CENP-A^{Cnp1} over-expression strain, the control strain, and purified sheared input (IN) DNA from both strains. Coverage was normalised by millions of reads mapped for all data sets so that total coverage was the same between data sets since the number of reads sequenced is unrelated to enrichment.

A statistic, $R_{IN,}$ measuring the difference in sequencing biases between the two strains, was calculated by dividing the coverage of the IN DNA from the CENP-A^{Cnp1} over-expression strain by the coverage of the IN DNA from the wild-type CENP-A^{Cnp1} expression strain on a per base level. All CENP-A^{Cnp1} ChIP samples are expected to have similar biases so normalising relative to the wild-type CENP-A^{Cnp1} ChIP strain should also take into account sequencing and extraction biases, but the additional normalisation of the sheared IN data in a similar way directly tests for differences in bias between the two experimental groups. A second statistic, $R_{IP,}$ measuring ectopic CENP-A^{Cnp1} incorporation, was calculated by taking the ratio of coverage between the CENP-A^{Cnp1} over-expression strain and the wild-type CENP-A^{Cnp1} expression strain at per base resolution. This gives a measure of CENP-A^{Cnp1} occupancy following over-expression relative to its normal occupancy.

The distributions of R_{IN} and R_{IP} values are shown in Figure 4.2. Log2 R_{IN} values are tightly distributed around a mean of zero, consistent with similar sequencing biases between the two experimental groups. In contrast, log2 $R_{IP,}$ values were more widely distributed. There is a higher frequency of positive R_{IP} values (especially above a log2 value of 1) compared to the R_{IN} distribution that reflects higher overall read coverage from cells that over-expressed CENP-A^{Cnp1}. This finding is consistent with ectopic CENP-A^{Cnp1} incorporation at a subset of sites in the genome. There is also a higher frequency of negative R_{IP} values (especially below a log2 value of -1) compared to the R_{IN} distribution that reflects lower read coverage obtained from cells with CENP-A^{Cnp1} over-expressed. Since the number of sequenced reads in a

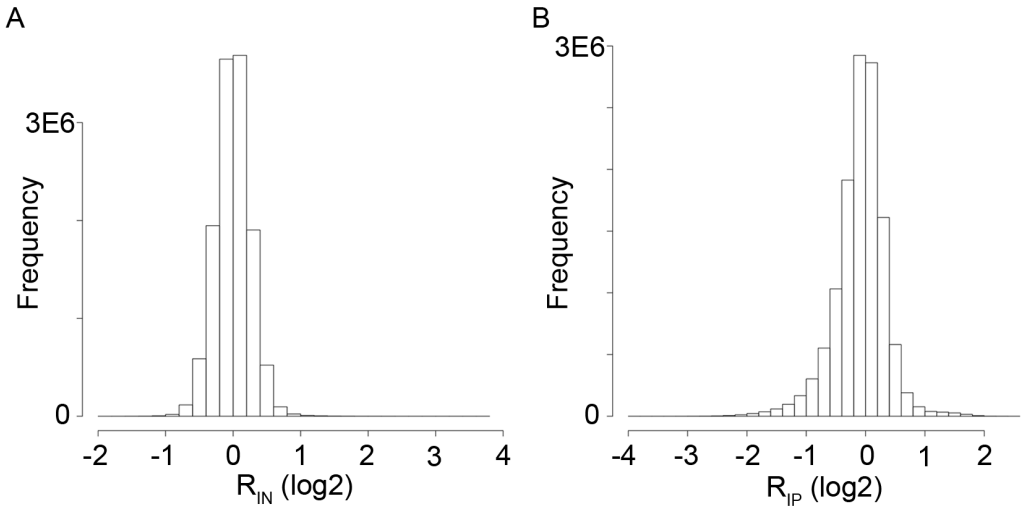


Figure 4.2. Test statistics show expected distributions. (A) Log2 R_{IN} (ratio between IN coverage of CENP-A^{Cnp1} over-expression strain and wild-type CENP-A^{Cnp1} expression strain) are tightly distributed around zero, indicating that most bases have highly similar IN coverage between the two strains. (B) Log2 R_{IP} (ratio between IP coverage of CENP-A^{Cnp1} over-expression strain and wild-type CENP-A^{Cnp1} expression strain) are more broadly distributed with a greater proportion of large values that reflect ectopic CENP-A^{Cnp1} incorporation and a longer negative tail due to the concomitant decrease in reads available to cover the rest of the genome.

ChIP-seq experiment is not related to the enrichment of the protein of interest, the coverage is scaled relative to the number of reads mapping to the genome to allow comparisons to be made between samples. A greater number of reads mapping to sites of ectopic CENP-A^{Cnp1} incorporation means that a smaller proportion of reads will map to the rest of the genome, which would cause more negative R_{IP} values at these other sites. It is also possible that the more negative R_{IP} values may be due to an unknown change in chromatin structure that makes these sequences more difficult to extract following CENP-A^{Cnp1} over-expression.

R_{IN} gives a measure of the noise and sequencing bias between the data sets so was used to calculate p-value significance thresholds, defining what constituted significantly higher and lower enrichment following CENP-A^{Cnp1} over-expression. The top 1% highest R_{IN} values defined the cut-off for positions with significant higher enrichment and likewise the bottom 1% lowest R_{IN} values determined the cut-off for significantly lower enrichment following over-expression. R_{IN} values were distributed around a mean of ~1, reflecting equal coverage between samples, with 1% of sites defined as significantly more enriched and 1% defined as significantly less enriched, with cut-offs of 1.528 and 0.654 respectively. As such, 1% of positions in the IN data were by definition considered to have higher or lower levels of enrichment based on their R_{IN} values even though IN coverage is expected to be the same between experimental groups.

R_{IN} values are shown across the length of all chromosomes in Figure 4.3. R_{IN} values more extreme than the upper and lower significance thresholds are shown in green and red, respectively. These are 1% of the genomic positions, by definition, and represent over- and under-representation of IN coverage in the CENP-A^{Cnp1} over-expression samples relative to the wild-type CENP-A^{Cnp1} expression samples. The notable clusters of significant red peaks correspond to the centromere central domains and suggest a change in chromatin structure at these loci that makes these sequences more difficult to extract in cells over-expressing CENP-A^{Cnp1}. It is possible that this is due to changes in CENP-A^{Cnp1} and H3 nucleosome occupancy or interactions with the kinetochore, but this remains to be tested. The green peaks of higher R_{IN} values are distributed across the genome and do not show any clear clustering at specific sites.

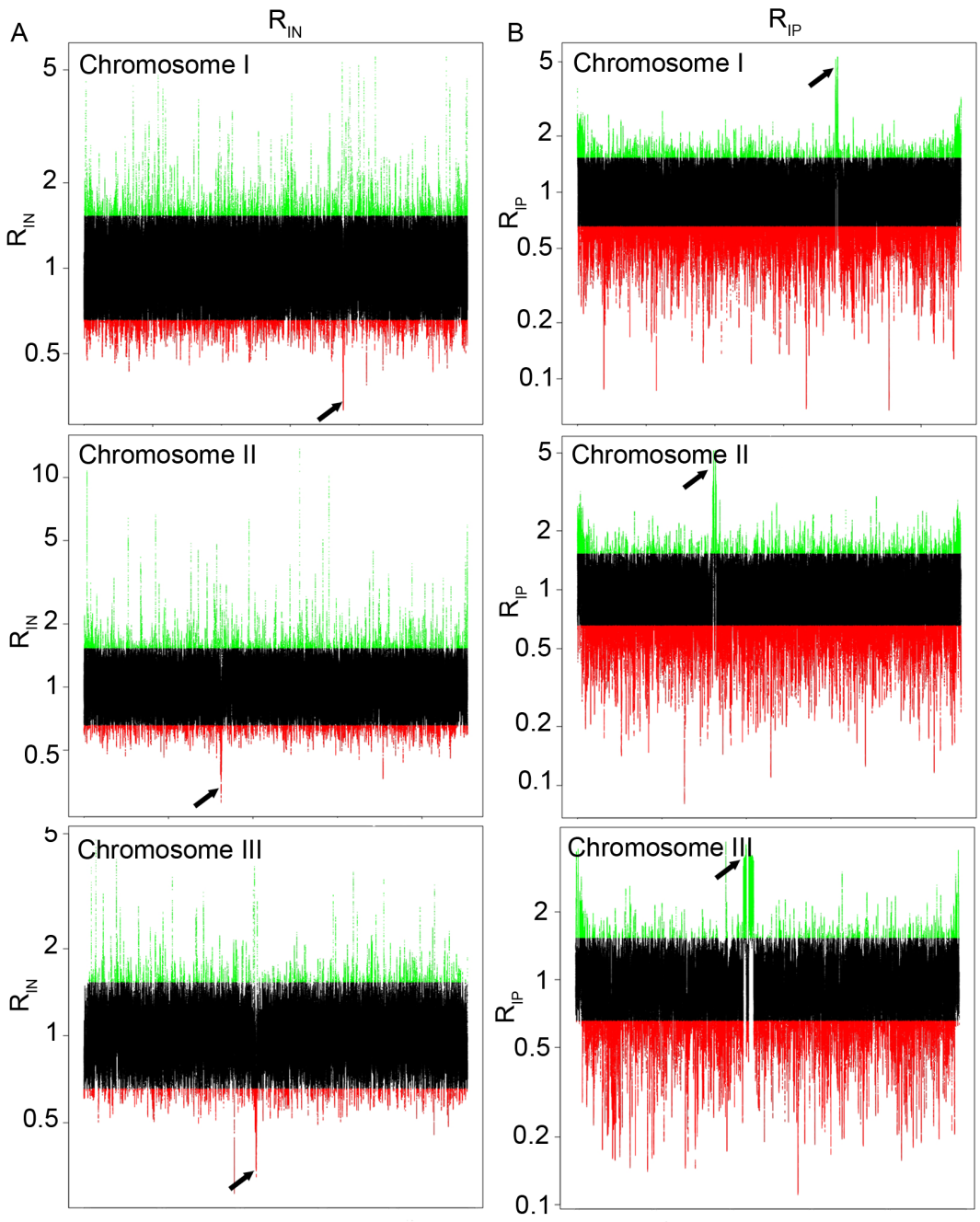


Figure 4.3. Ectopic CENP-A^{Cnp1} incorporation primarily occurs at the centromere and subtelomeres. (A) R_{IN} values are shown plotted across the length of the fission yeast chromosomes at single basepair resolution. Arrows indicated centromeres. A p-value significance threshold was calculated so that the highest and lowest 1% of sites were considered to have significantly higher or lower coverage in the CENP-A^{Cnp1} over-expressed strain relative to the wild-type CENP-A^{Cnp1} expression strain. These sites are indicated in green and red for higher and lower coverage, respectively. (B) The p-value thresholds were applied to the R_{IP} values across the three chromosomes. There is an enrichment of significant ectopic incorporation at the subtelomeres and centromeres.

The p-value significance thresholds based on R_{IN} values were applied to the R_{IP} values to define significant changes in CENP-A^{Cnp1} enrichment following over-expression. On all three chromosomes there are clear clusters of significantly highly enriched sites around the centromeres and over subtelomeric regions, as indicated by the green peaks. This increase over specific regions is consistent with ectopic CENP-A^{Cnp1} incorporation at these loci (Figure 4.3). Enrichment is higher at the centromere than the subtelomeric regions. The R_{IP} values at these loci, especially the subtelomeric regions, are not as high as some of the observed R_{IN} values. The distributions shown in Figure 4.2 reveal that high R_{IN} values are due to a very small number of bases with extreme values, while high R_{IP} values are due to a much larger number of bases with moderately high values. This suggests that the high R_{IN} values are due to noise, while high R_{IP} values are due to ectopic CENP-A^{Cnp1} incorporation.

To determine whether there are other regions of broad enrichment, the average R_{IP} value and number of significant bases was calculated in 10 kb non-overlapping bins across the genome. It is clear that sites near centromeres and subtelomeres have higher R_{IP} values and more bases with significant ectopic enrichment, while no other areas of the genome show a similar pattern (Figure 4.4). Both these measures indicate that there is higher enrichment near the centromere, meaning that this is the primary site of ectopic CENP-A^{Cnp1} incorporation. However, ectopic incorporation can also be investigated at a finer spatial resolution and therefore the analyses were repeated using 1 kb non-overlapping bins across the genome. The highest levels of enrichment and significant bases still occur at the outer-repeats and subtelomeric regions, but there are additional bins across the genome that show evidence of ectopic CENP-A^{Cnp1} incorporation. These ectopic incorporation sites occur on a smaller spatial scale since they were not detected when using broader 10 kb bins that would include nearby regions where ectopic incorporation did not occur. Nevertheless, this finding provides evidence for additional ectopic CENP-A^{Cnp1} incorporation at additional locations across the genome and that it is not restricted to the outer-repeats or subtelomeric regions. This is discussed further below.

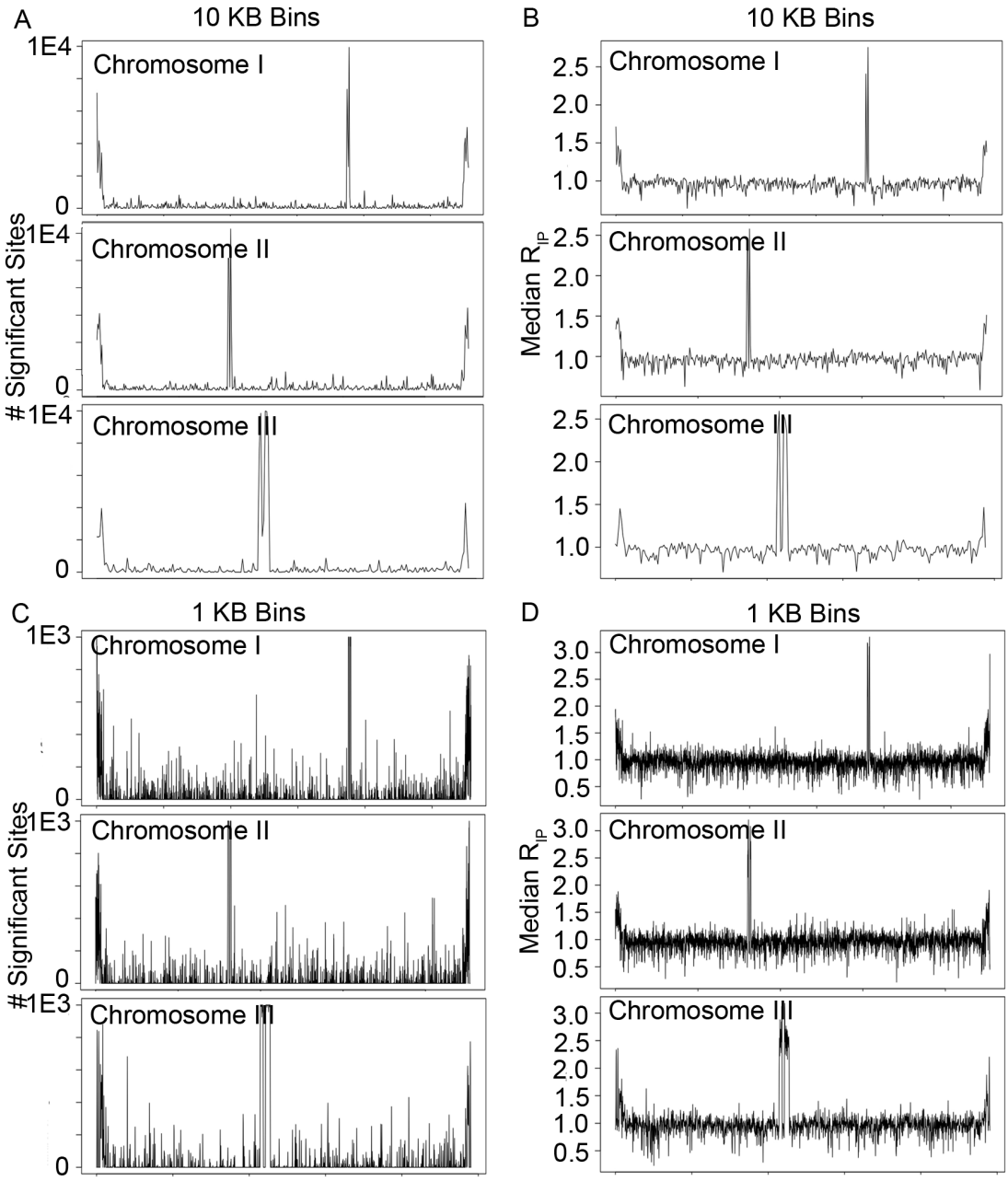


Figure 4.4. There are small regions of ectopic CENP-A^{Cnp1} incorporation across the genome. (A) The number of sites above the p-value significance threshold that indicates statistically significant ectopic CENP-A^{Cnp1} incorporation and (B) the median R_{IP} values calculated for 10 kb non-overlapping bins genome-wide are shown for all three chromosomes. The centromeres and subtelmeric regions are identified as having an enrichment of significant bases and higher R_{IP} values. No other sites across the genome show evidence of wide-spread ectopic CENP-A^{Cnp1} incorporation. However, when (C) the number of significant sites and (D) the median R_{IP} values were calculated for 1 kb non-overlapping bins there is evidence of limited ectopic CENP-A^{Cnp1} incorporation at other sites across the genome, though the centromere and subtelomeres still have the highest enrichment.

4.2.3 Ectopic incorporation at centromeres is primarily at the outer repeats

The highest levels of significant ectopic CENP-A^{Cnp1} incorporation were detected at centromeres. Ectopic CENP-A^{Cnp1} incorporation was limited to the centromeric outer-repeats that surround the central domain at each centromere (Figure 4.5). Coverage is plotted across the centromere of chromosome II since it is unique while centromeres on chromosomes I and III share a region of homology, but all analyses include all three centromeres. There was no detectable ectopic CENP-A^{Cnp1} incorporation within the central core or over the inner-repeats, even on the heterochromatic portion of the inner-repeats distal to the tRNA boundary elements. In partial contrast to the ChIP-qPCR analysis (Figure 4.1), CENP-A^{Cnp1} was not higher in the CENP-A^{Cnp1} over-expression samples, though (as mentioned above) it is not clear whether the ChIP-qPCR results support increased CENP-A^{Cnp1} occupancy due to the decrease in central domain chromatin extraction following CENP-A^{Cnp1} over-expression. ChIP-seq analysis is expected to have less power than ChIP-qPCR to detect this difference, due to the lower dynamic range and constraint on the total number of reads sequenced, so it is not surprising that no ectopic CENP-A^{Cnp1} incorporation is detected within the central domain. Thus, it is possible that following CENP-A^{Cnp1} over-expression there is some additional CENP-A^{Cnp1} incorporation within the central domain, as well as at other regions in the genome, but this cannot be detected using the current ChIP-seq methodology without a spike in chromatin control.

4.2.4 Increased CENP-A^{Cnp1} expression shifts CENP-A^{Cnp1} occupancy within the central domain towards sites of predicted nucleosome occupancy

Ectopic CENP-A^{Cnp1} incorporation could not be detected within the central domain following CENP-A^{Cnp1} over-expression despite potential increased occupancy of CENP-A^{Cnp1} detected by ChIP-qPCR. However, CENP-A^{Cnp1} occupancy in cells over-expressing CENP-A^{Cnp1} show a different occupancy pattern within the central domain compared to cells with wild-type CENP-A^{Cnp1} expression. There are clear peaks and troughs where CENP-A^{Cnp1} coverage is higher or lower in CENP-A^{Cnp1} over-expressed cells compared to control cells, though the changes in occupancy are relatively small and not significant. This is of particular interest considering the pattern of highly positioned CENP-A^{Cnp1} observed within the central domains (Chapter 3). Patterns of coverage within the central domain were analysed in greater detail to determine what factors influence this pattern of peaks.

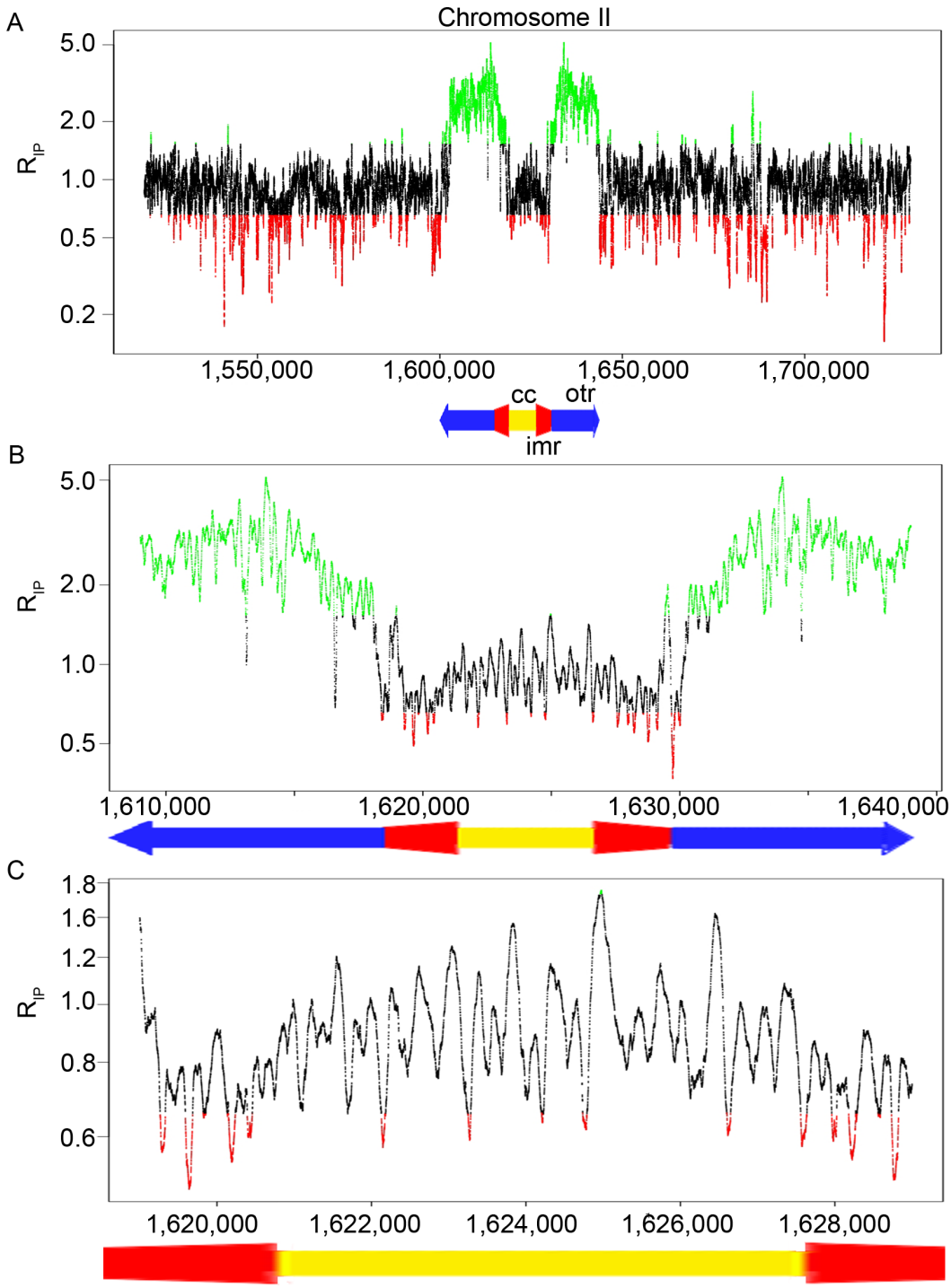


Figure 4.5. Ectopic CENP-A^{Cnp1} incorporation is specific to the outer-repeats. The chromosome II centromere is shown with (A) 100 kb, (B) 10 kb, (C) and 0 kb of flanking sequences. Green and red represent significantly higher or lower coverage in the CENP-A^{Cnp1} over-expressed strain relative to the wild-type CENP-A^{Cnp1} expression strain, respectively. Significant ectopic CENP-A^{Cnp1} incorporation (green) is limied to the outer-repeats (otr) with no significant ectopic incorporation detected on the central core (cc) or inner-repeats (imr).

Across the central domains R_{IP} values did not provide evidence of ectopic CENP-A^{Cnp1} incorporation at additional positions, but a pattern of periodic peaks was observed (Figure 4.6A). Control samples displayed high enrichment of CENP-A^{Cnp1} across the central domain with peaks corresponding to positions of highest enrichment (Figure 4.6B). Due to the described technical limitations (Chapter 3), CENP-A^{Cnp1} only shows a slight enrichment at the peaks and it is difficult to rigorously define peak positions. Therefore, nucleosome positioning cannot be defined at a high resolution and the shifts in occupancy detected between the two groups represent only subtle changes. Cells over-expressing CENP-A^{Cnp1} had a high enrichment of CENP-A^{Cnp1} across the central domain (Figure 4.6C). This enrichment again showed a high general enrichment with peaks corresponding to positions of highest enrichment, but there is a clear shift towards more precisely defined peaks. This shift corresponds to the peaks observed in the R_{IP} values across the central domain.

R_{IP} values were only weakly correlated with wild-type CENP-A^{Cnp1} occupancy (τ=0.183; Figure 4.6D), indicating that the peaks did not simply represent higher enrichment at the peaks already present. Rather, R_{IP} values were moderately correlated with the predicted nucleosome occupancy calculated using the Segal algorithm (τ=0.382; Figure 4.6E). Peaks of higher R_{IP} values are associated with predicted nucleosome occupied sites. Other new peaks correspond to regions within the central domain that are not predicted to be occupied by nucleosomes and may form gaps, which is surprising considering that these regions are more AT-rich and CENP-A^{Cnp1} does not normally occupy all predicted sites (Chapter 3). However, confirmation of these analyses will need nucleosome positions to be mapped at higher accuracy by other methods such as digesting chromatin with MNase to single nucleosome resolution or ChIP-exo to map positions of nucleosome-DNA cross-linking.

4.2.5 Ectopic CENP-A^{Cnp1} incorporation occurs at sites where stable neocentromeres are known to form

Formation a neocentromere can be induced in fission yeast by the deletion of a canonical centromere from a chromosome (Ishii et al., 2008). When canonical centromere I is deleted a neocentromere can form in the subtelomeric region near either telomere of chromosome I (Ishii et al., 2008). Therefore, it was of interest to

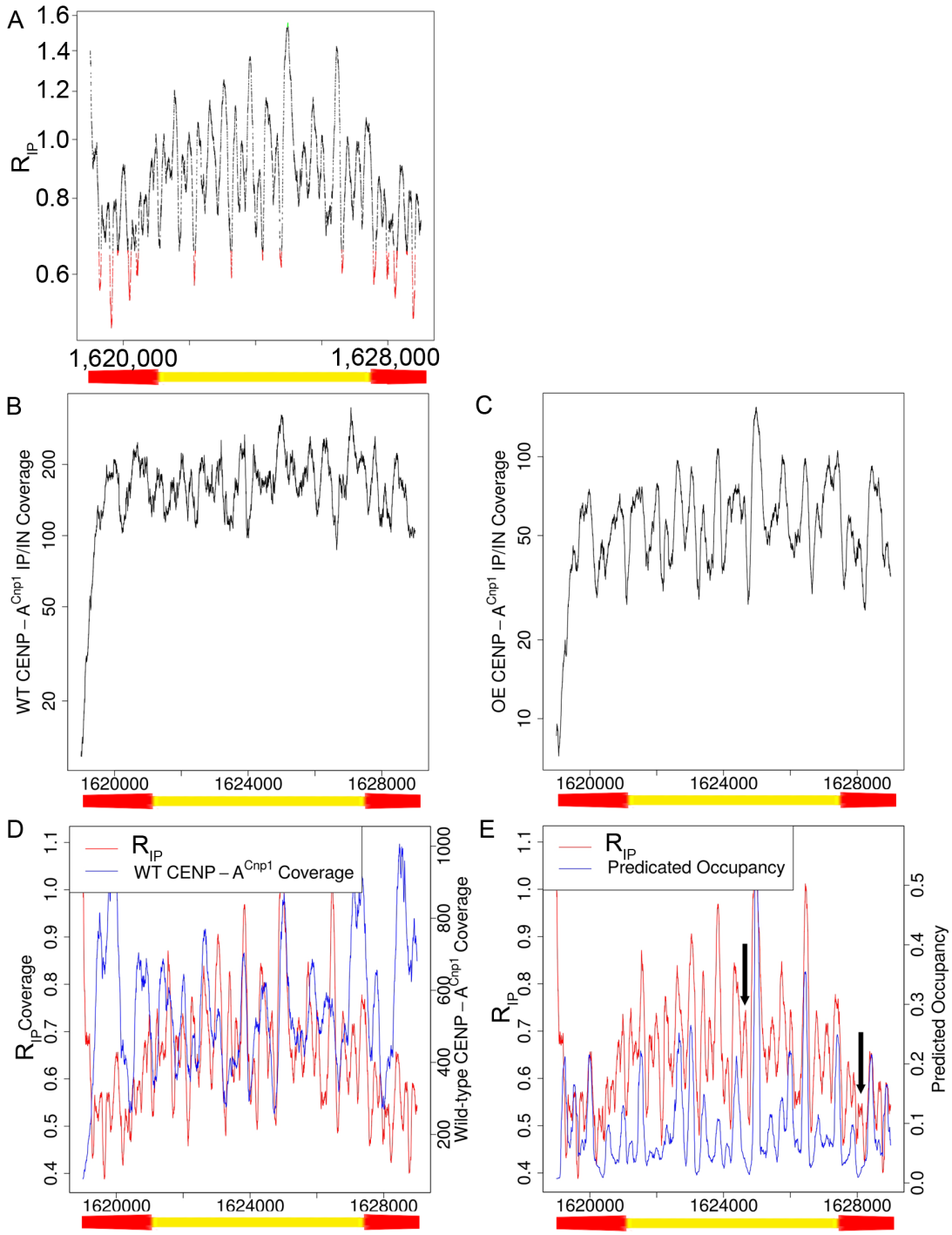


Figure 4.6. Ectopic incorporation leads to more positioned CENP-A^{Cnp1} occupancy. (A) CENP-A^{Cnp1} over-expression leads to certain sites within central domain II having slightly higher coverage. (B) CENP-A^{Cnp1} coverage in wild-type cells showed high enrichment across the central domains. (C) CENP-A^{Cnp1} coverage in cells over-expressing CENP-A^{Cnp1} had more defined peaks. (D) The shift in occupancy following over-expression is weakly correlated with the wild-type CENP-A^{Cnp1} occupancy (τ=01.83). (E) The shift in occupancy following over-expression is moderately correlated with predicted nucleosome occupancy (τ=0.382), potentially with additional peaks (examples indicated by arrows).

determine whether those regions that are capable of neocentromere formation correspond to regions of ectopic CENP-A^{Cnp1} incorporation following CENP-A^{Cnp1} over-expression. Previously sequenced CENP-A^{Cnp1} ChIP-seq data for neocentromere carrying strains with wild-type CENP-A^{Cnp1} expression levels (Chapter 3) were compared with ectopic CENP-A^{Cnp1} incorporation due to CENP-A^{Cnp1} over-expression in strains with canonical centromeres.

Neocentromere cd39 is located near the left telomere of chromosome I (Ishii et al., 2008). ChIP-seq of CENP-A^{Cnp1} in a strain carrying the cd39 neocentromere showed high enrichment of CENP-A^{Cnp1} at the known neocentromere site near the left telomere of chromosome I (Figure 4.7; data from H. Berger, Allshire lab). In comparison, the strain with over-expressed CENP-A^{Cnp1} showed significant ectopic incorporation over a broad domain in the subtelomeric region (Figure 4.7). There was significant ectopic incorporation of CENP-A^{Cnp1} following CENP-A^{Cnp1} over-expression (OE) at the site of neocentromere cd39, but ectopic incorporation is not limited to the neocentromere region. Neocentromere cd39 was shown to be located near the telomere ~70 kb from the chromosome end (Ishii et al., 2008). Consistent with this, CENP-A^{Cnp1} is enriched over a domain that lies ~70-90 kb from the left telomere of chromosome I. In contrast, ectopic CENP-A^{Cnp1} incorporation following over-expression leads to enrichment from the telomere extending across the neocentromere site (~0-90 kb). However, the CENP-A^{Cnp1} signal at neocentromere cd39 and in OE CENP-A cells decline at approximately the same position internal to the telomere, but ectopic incorporation may extend internally to the neocentromere position. This boundary corresponds to the limit of CENP-A^{Cnp1} incorporation is not associated with changes in AT-content and occurs near the *str3* gene (located at 88,201-90,300 on the reverse strand) or the region downstream.

Neocentromere cd60 was shown to be located in the subtelomeric region near the right telomere of chromosome I (Ishii et al., 2008). Our ChIP-seq data for CENP-A^{Cnp1} on the cd60 strain showed high enrichment of CENP-A^{Cnp1} near the right telomere of chromosome I (data from H. Berger, Allshire lab). In comparison, the strain with over-expressed CENP-A^{Cnp1} showed significant ectopic incorporation over a broad domain in this subtelomeric region (Figure 4.8). Similar to neocentromere cd39, there was significant ectopic incorporation of CENP-A^{Cnp1} in cells over-expressing CENP-A^{Cnp1} in the region where neocentromere cd60 formed,

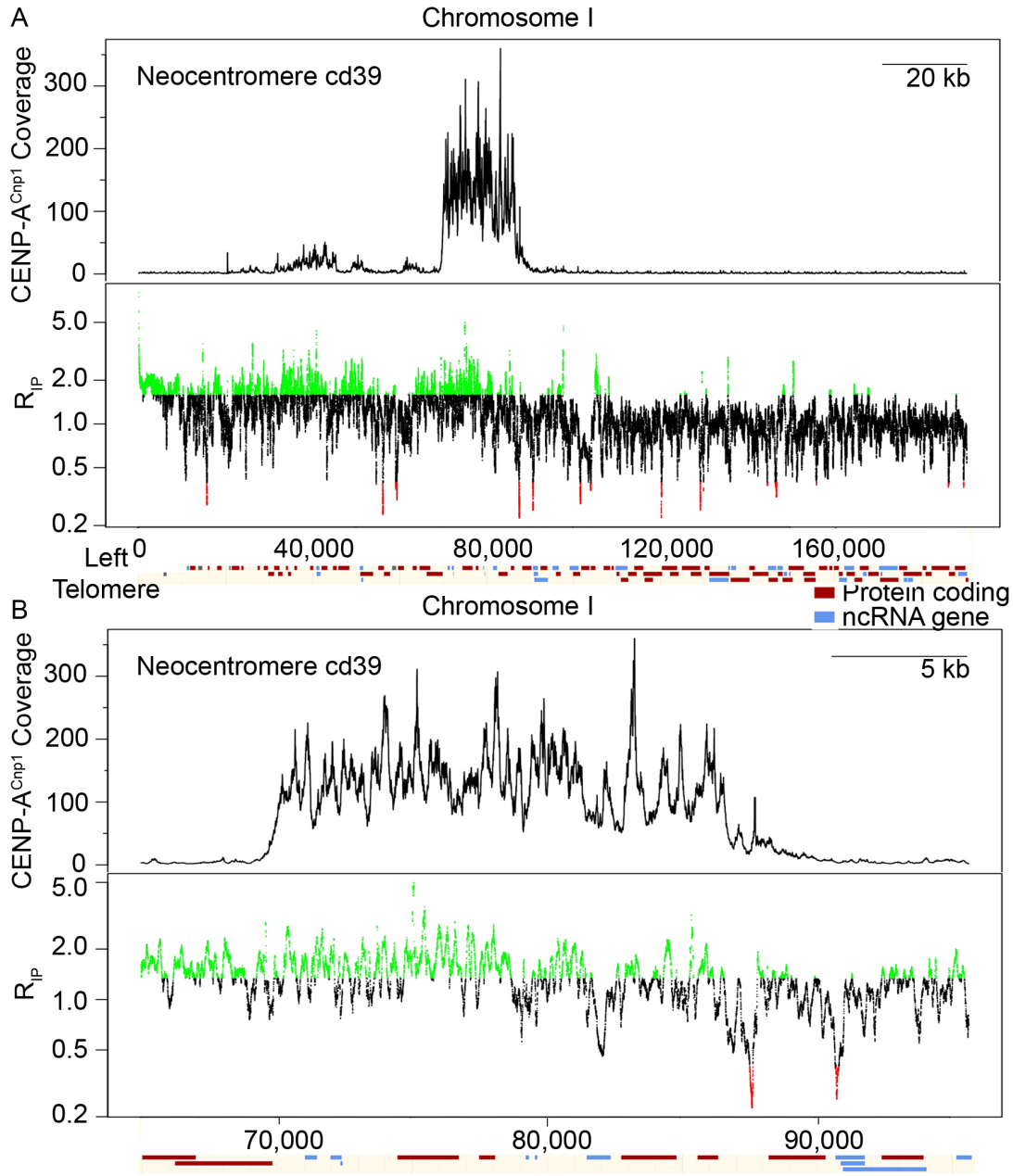


Figure 4.7. Ectopic CENP-A^{Cnp1} incorporation occurs at the site neocentromere cd39 forms. CENP-A^{Cnp1} ChIP-Seq coverage in a cd39 neocentromere strain with wild-type CENP-A^{Cnp1} expression (top; data from H. Berger; shown in Chapter 3) and ectopic incorporation following CENP-A^{Cnp1} over-expression (R_{IP}; bottom) is shown at neocentromere cd39 with (A) 100 kb and (B) 10 kb flanking sequences. Green and red represent significantly higher or lower coverage in the CENP-A^{Cnp1} over-expressed strain relative to the wild-type CENP-A^{Cnp1} expression strain, respectively. There is significant ectopic incorporation following CENP-A^{Cnp1} over-expression at the site of neocentromere formation, but ectopic incorporation is not limited to this site. Ectopic incorporation also occurs from the telomere and extending across the length of the regions where neocentromere cd39 forms.

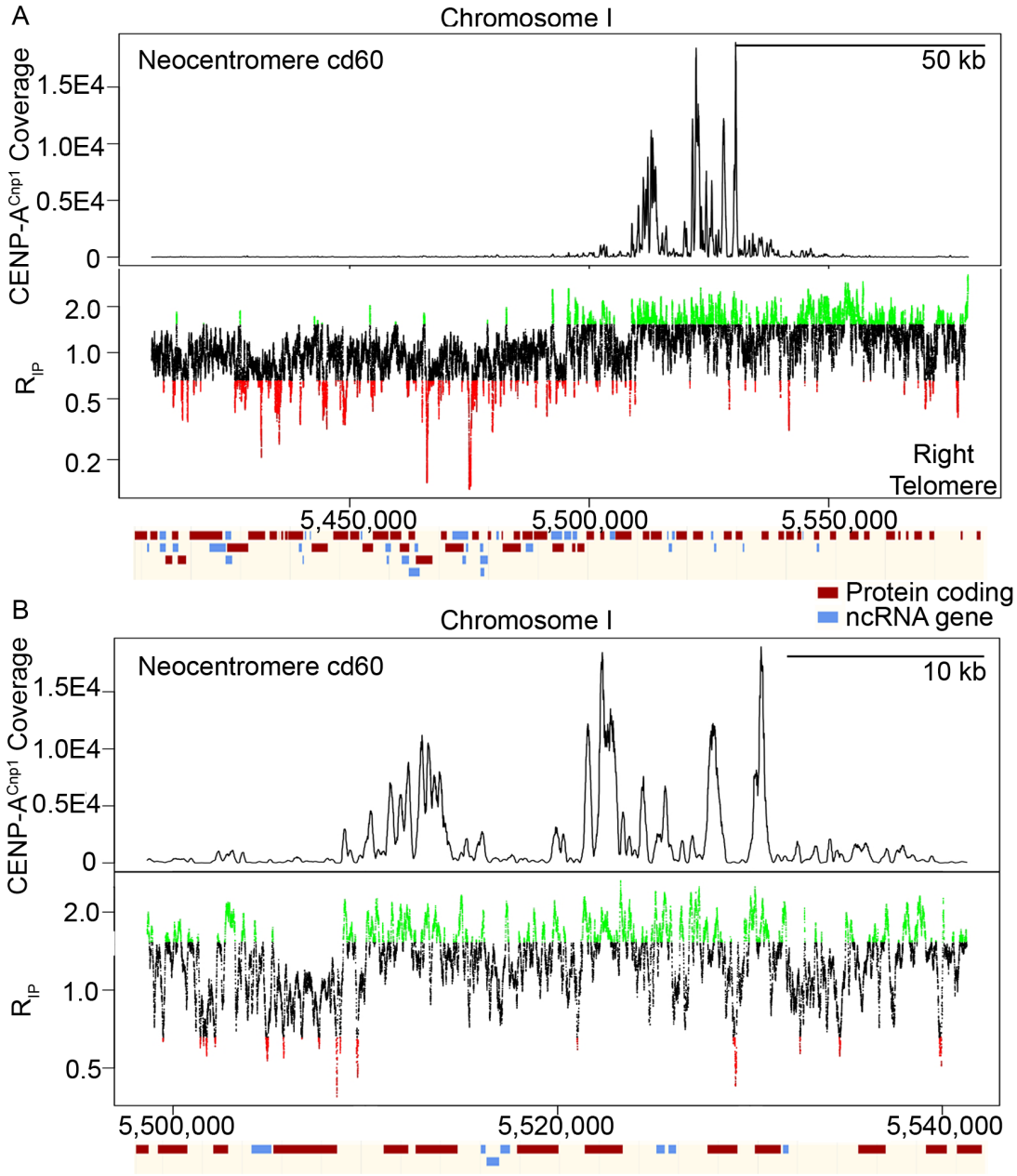


Figure 4.8. Ectopic CENP-A^{Cnp1} incorporation occurs at the site neocentromeres cd60 forms. CENP-A^{Cnp1} ChIP-seq coverage in a cd39 neocentromere strain with wild-type CENP-A^{Cnp1} expression levels (data from H. Berger; shown in Chapter 3) and ectopic incorporation following CENP-A^{Cnp1} over-expression (R_{IP}) is shown at neocentromere cd60 with (A) 100 kb and (B) 10 kb flanking sequences. Green and red represent significantly higher or lower coverage in the CENP-A^{Cnp1} over-expressed strain relative to the wild-type CENP-A^{Cnp1} expression strain, respectively. There is significant ectopic incorporation following CENP-A over-expression at the site of neocentromere formation, but ectopic incorporation is not limited to this site. Ectopic incorporation also occurs from the telomere and extending across neocentromere cd60.

but ectopic incorporation was not limited to the neocentromere region. Neocentromere cd60 is known to reside ~50 kb from the right telomere of chromosome I and CENP-A^{Cnp1} in cd60 cells is enriched over a domain positioned ~50-70 kb from right end of chromosome 1. However, in OE CENP-A^{Cnp1} cells a broad region of ectopic incorporation was indicated by CENP-A^{Cnp1} enrichment that extended from the telomere over the neocentromere cd60 region and possibly internal into the chromosome arm (~0-90 kb).

4.2.6 Ectopic CENP-A^{Cnp1} incorporation is limited at an internal site capable of neocentromere formation

When the centromere was deleted from chromosome III it was observed that neocentromere formation was rare, likely due to the presence of rDNA repeats adjacent to both telomeres (Ogiyama et al., 2013). Unlike the formation of subtelomeric neocentromeres that followed deletion of the centromere of chromosome I (Ishii et al., 2008), subtelomeric neocentromeres were not observed on chromosome III following deletion of the chromosome III centromere and instead neocentromeres were formed at internal sites (Ogiyama et al., 2013). Two internal neocentromeres were characterized, one (cd389) located on the side of the rDNA repeat array lying internal on the chromosome (~1Mb from the telomere) and one (cd385) located at an internal euchromatic locus (at least 500 kb from either telomere). Neocentromeres cd385 and cd389 are unstable and show segregation defects but can be stabilised by chromosome fusion or rearrangements that bring the neocentromeres closer to the telomeres. Stable neocentromeres on chromosome I were formed near the telomeres so their locations may be determined primarily by chromatin context, such as associated telomeric heterochromatin, rather than any sequence features. In contrast, the loci involved in forming these unstable neocentromeres on chromosome III were not positioned near centromeres or telomeres or other unusual chromatin. Thus these non-telomeric neocentromeres may be determined by underlying sequence features at these locations. To determine if these same features promote ectopic CENP-A^{Cnp1} incorporation in these regions CENP-A^{Cnp1} levels was analysed over these potential neocentromere regions in CENP-A^{Cnp1} over-expressing cells.

No widespread significant ectopic CENP-A^{Cnp1} incorporation in CENP-A^{Cnp1} over-expressing cells was detected in the region where neocentromere cd385 can form.

However, a small segment within the region covered by neocentromere cd385 displays significant ectopic CENP-A^{Cnp1} incorporation (Figure 4.9). It is possible that this segment (s385) may seed CENP-A^{Cnp1} incorporation during neocentromere formation and that additional CENP-A^{Cnp1} may then be recruited into nearby chromatin. The s385 segment is small (~0.5 kb) but it displays significant ectopic CENP-A^{Cnp1} incorporation. However, given the small size of this segment it is unclear whether it has any functional significance with respect to neocentromere formation.

Neocentromere cd389 formed on the internal side of the rDNA repeats near the left telomere of chromosome III (Ogiyama et al., 2013). In cells with canonical centromeres that overexpress CENP-A^{Cnp1} widespread significant ectopic incorporation of CENP-A^{Cnp1} was found to overlap with the cd389 region. Neocentromere cd389 is considered an internal on the chromosome since it is located on the internal side of the rDNA repeat array, but the rDNA repeats show broad ectopic CENP-A^{Cnp1} incorporation that extends across the site where neocentromere cd389 formed. It cannot be determined which rDNA repeats show ectopic CENP-A^{Cnp1} incorporation, but the significant enrichment over the region including cd389 suggests that there is likely ectopic incorporation across the length of the rDNA repeats. As such, ectopic incorporation in this region may be related to chromatin context due to proximity to the telomere or the rDNA repeats themselves rather than specific sequence features. Indeed, heterochromatin has also been shown to be associated with the rDNA repeats in fission yeast (Bjerling et al., 2004; Shankaranarayana et al., 2003).

4.2.7 Ectopic CENP-A^{Cnp1} incorporation occurs at sites enriched for asymmetric AT-rich sequences

It was of interest to determine what sequence features might lead to ectopic CENP-A^{Cnp1} incorporation at regions outside outer-repeats and subtelomeric regions. One possibility is that these regions share some sequence feature that somehow promotes ectopic CENP-A^{Cnp1} incorporation. These regions were therefore analysed to determine if they share any sequence motifs that might correlate with ectopic CENP-A^{Cnp1} incorporation.

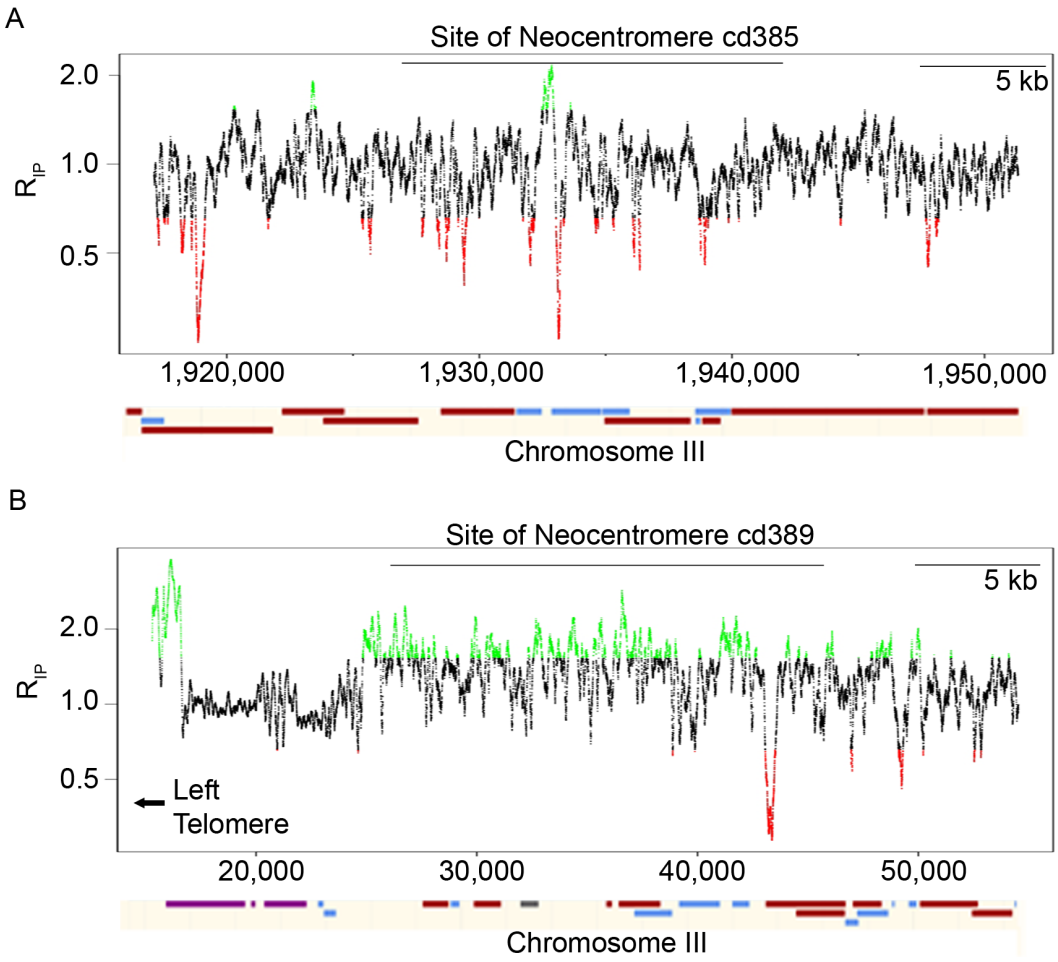


Figure 4.9. Ectopic CENP-A^{Cnp1} incorporation is limited at the site an internal neocentromere forms. Ectopic CENP-A incorporation following over-expression is shown at the regions where unstable internal neocentromeres (Ogiyama et al. 2013) cd385 (A) and cd389 (B) form on chromosome III. Green and red represent significantly higher or lower coverage in the CENP-A^{Cnp1} over-expressed strain relative to the wild-type CENP-A^{Cnp1} expression strain, respectively. At the locus where internal unstable neocentromere cd385 forms there is minimal ectopic incorporation, although there is one site of significant ectopic incorporation within the region (~500 bp). It is not clear whether this small amont of ectopic incorporation is relevant to neocentromere formation. In contrast, at the locus where neocentromere cd389 forms there is broad ectopic CENP-A^{Cnp1} incorporation following over-expression, likely due to its proximity to the rDNA repeats and telomeres.

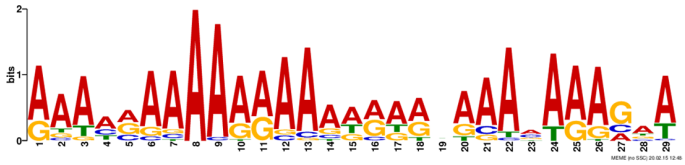
All 1 kb bins with significantly positive R_{IP} values that were not in within the annotated centromeres or within 100 kb of the telomeres were identified. Only bins with at least 300 significantly positive bases were used due to computational limitations. This number was chosen since the maximum fragment size mapped was 250 bp so that any significant enrichment would be due to multiple non-identical reads. 43 bins fit these criteria. Shared motifs present in these bins were identified using MEME (Bailey and Elkan, 1994). The most significant motif identified was a highly asymmetrical AT-rich motif of width 29 bp that had an E-value of 5.6e-21 (Figure 4.10). This motif was present in 29 of the 43 bins. Two other significantly enriched motifs were identified. The first motif is an asymmetrical AT-rich motif present in 17 bins (E-value=2.4e-5; Figure 4.11). The second motif is an asymmetrical AT- and asymmetrical CG-rich motif present in 17 bins (E-value= 9.3e-3; Figure 4.12).

The asymmetrical AT-rich motifs identified at these internal sites of ectopic CENP-A^{Cnp1} incorporation are similar to the motifs that were observed in the gaps between CENP-A^{Cnp1} nucleosomes at canonical centromeres (Chapter 3). Thus, these asymmetrical AT-rich motifs are candidates for sequences that influencing ectopic CENP-A^{Cnp1} incorporation. AT-rich sequences generally exclude nucleosomes from assembly and the analysis applied used 1 kb windows so it is likely that ectopic CENP-A^{Cnp1} incorporation occurs on the surrounding region and not over the AT-rich sequence itself. Further analyses on where ectopic incorporation occurs relative to these sequence features and what the normal pattern of canonical nucleosomes is in the regions are being undertaken.

5.2.8 Ectopic CENP-A^{Cnp1} incorporation does not occur at replication origins

The asymmetrical AT-rich motif present at bins of significant ectopic CENP-A^{Cnp1} incorporation is intriguing, but it is not clear how these two are related. The origin recognition complex (ORC) localises to replication origins in fission yeast via the multiple AT-hooks in the Orc4 protein (Chuang and Kelly, 1999; Kong and DePamphilis, 2001; Lee et al., 2001). Interestingly, the Orc4 AT-hooks bind to

A

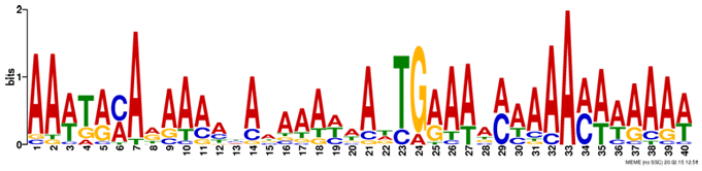


B

Name	Strand	Start	p-value	Sites
43	+	614	2.01e-12	AGAATCTTAA AAAAAAAAAAAAAAAGGACAAACAAGGAA ACCTTAAGGG
17	-	699	2.01e-12	AAATTTTTAA AAAAAAAAAAAAAAAAAAAAAAAAAAGAA TAGAGAAAAG
19	+	908	8.13e-12	CACAGACAAC AAAAGAAAAAAAAAAGAGTACACAAAGTA AACAATGGTG
40	-	698	7.14e-10	TGAATGAGAG AAAAGAAAAGAAAAGGAATGATAAGAGAA GGAATGAAAA
27	+	382	9.43e-10	TCCACTTGTT AAAACAAAAAAAAAACGAAAAACAGAAAA ATAAACCCTC
35	-	266	3.02e-09	GAAATAAAAT AAAACGAAAGAAAATAAATAAACAAAAGA ATACTCGAAA
16	-	761	7.73e-09	AAGTTTCGTC AAGCGAAAAAGAAAGAAGGAAAATAAAAA TTTTCTTCCA
7	-	478	9.64e-09	TTCATTTAGT AAAAAAAAAAAAAAAAATTGACTAAGCA ATACGAAATA
33	+	284	1.08e-08	AGGTGTAGTA AAACAAAAAAAACTAAGGTAAAGAGAGAA TTGCTTGAAT
39	-	31	1.65e-08	TGCAGCAAAG ACACGAAAAAGAACAGAGGAAACTGAAAA TTGGGCAAGC
8	-	681	1.65e-08	TTACATTCAA GGATGAAAAAGCAAGAGAGGAAGAAAGAA ACTCACCTTA
1	-	270	2.24e-08	TTTTCGGACA AATTCAAAAAACAAAAAACAAAAAAACCA TTCTGTTCAT
37	+	382	2.73e-08	TCTAAGCAAC GAAAGAGAAAAAAGAATAGAAAAAAACGT CTATTAACAA
20	-	832	2.73e-08	TTCTTCTTTC AAATAAAACAAAAATGAACAAAGAAAGTT ACCCTCTATT
5	-	813	2.73e-08	AACCAAACGT AAACGATAAAAAAGTAAGCAAACAAGGGT TTAGTTGAAG
42	-	763	3.32e-08	AAACGAGACG AAAGGAAAAGAAAATAAAATAAAAGGGAA AAAAAATAAA
12	+	589	4.42e-08	AAGTGTTTTG GTACAAAAAAGAGAAGAGAAAACAAAGCT AGTATTAGAT
2	-	435	4.42e-08	TTTCGGTAGC AAAAGCAAAGGGAATGAAAAAATAAAGCA TTAAACGGAT
41	+	441	5.84e-08	TGGTTGAGTT AATACAAAAAAAAGGAAATGAACAAGCTT TTTATTAGAA
3	+	933	1.64e-07	GAGCTGGGCG AAAAAAAAAGGAACAGGTTGCAAAAAGGC ACTATACATT
21	+	840	2.63e-07	GGATGCGTGC GAAAGAAAAAGAACGATGCGAAATGAATT GTAAAGTATG
26	+	688	3.30e-07	AAGAAAGAGA ATTTAAGAAAAAAAATGGATAAAGAGACGA AATTGCTGAA
23	+	587	4.12e-07	AGTTGATGAG GAAAAGAAAGGAAAGGTGAAGAGAATCAA TCAGCAGCAA
11	-	187	6.72e-07	ACATATTCAT GTATCAAAAAAAAGTCTACTAAAAGAGAT TTTAACAGTA
32	-	187	7.69e-07	CTACAGTACA AAATACCAATAAAAAGTGTGAATAAAGAA CATTTAAATC
34	+	725	1.06e-06	AAAATTATTG AGTCCGAAAAAAAATAATGAATAAAAGTA GATCCGACAA
4	-	907	1.21e-06	TTGATATGGG GGAAAACAAGAGCAAAAACACAAAAAAAA ACAATCTTTG
22	+	824	1.28e-06	GAGCTAACAA GAGCGGAAAAGAATAGAGAGGCAAAGGAA GAACACTATG
15	+	746	4.39e-06	ATTTGGTAAG ACTAAAAAAAGAAGACTTTACATTAACAA ATTATTTAGG

Figure 4.10. Ectopic CENP-A^{Cnp1} incorporation occurs at regions enriched for asymmetric AT-rich sequences. (A) MEME (Bailey and Elkan 1994) identified a 29 bp asymmetric AT-rich motif enriched at 1 kb bins that show evidence of ectopic CENP-A^{Cnp1} incorporation. (B) This motif has an E-value of 5.6e-2 and is found in 29 of the 43 sequences used.

A

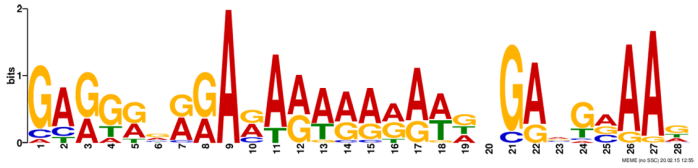


B

Name	Strand	Start	*p*-value	Sites
15	-	289	4.62e-15	TATCATGTGT AATTACAGAAAATAAAAACTATTGATAACAAAAAAAAAAA CCTTATTATT
14	+	780	4.62e-15	TATCATGTGT AATTACAGAAAATAAAAACTATTGATAACAAAAAAAAAAA CCTTATTATT
40	+	647	1.30e-11	GAACAAAGAC AAAGAAAACAAAAGAAAGCAATTGAAAACAAAACAAAACA ATTTTCATTC
42	-	672	1.47e-11	GAACCCTAAA AAATAAAAATAAAAGGGAAGAATGAAAACCAAAAATAAAA GTGAGGAATG
3	-	610	2.63e-11	ATATCGAAAT AAATGCAAAACAGATAAACAAATGGAAACCAAAAAACCGT TCAAGCATGT
43	+	522	6.74e-10	AAAGATGAAT GAATGAATGAATAATAAATAAATGGAACAAAAAAATGAAT GAATAAATTG
35	-	208	7.37e-10	AATAAAATAA AACGAAATAAAATAAAATTAAACGAAATAAACAAAAGAAT ACCTCTATTA
19	+	495	8.05e-10	TGTCGAGGCA AATGAAAAATATTCCCAAAACTTGGAAGAAAAACTAGAAT GGCGCTAACA
16	-	165	8.05e-10	AATACAGAAA CAAAACAGACGACACAATACAGTGGATAACAAACAAGAAA AATTGCCCCT
17	+	636	2.60e-09	GAAAGCAGTG AAATCCAAAAAAAAAAAAAAAATTAAATAAATCACTCAAAT CCAATCACTA
13	-	950	2.60e-09	ATTAATAAAA AAAAACAAGAATCCTAAAATATCAAATTCAAAAAAAAAGA ATTCGACACG
37	+	89	3.05e-09	TAAACAATGA AGATGCAAAACACAAGTATAGGTGAATGATGAACAAAAGA CTGCTTTGTT
10	-	539	3.84e-09	AGTAGTATAG AAATGCTGGACCAACCAGTTAATGACATATAAACATAAAA CATAAAAAAC
29	-	92	6.49e-09	TTTTACAAAG AATTACATAACGCAGTTTACCGTGGAAAAACAAATTAAAA ATTTGACCTA
25	-	573	6.49e-09	ACCTCAATAA ATTTAAAGAAACTAAAGATGATTGAAACCTCAACTCATAA AGCTATTCTC
2	+	858	9.96e-09	GAAAACAAGT AAATGTACATACCAAACAAAAGTGTAAAATAAAATAAAAC AAAAAGTTCT
18	-	120	1.40e-08	GGTATATGGA AATGAAAGCAGCGCGTTATTACCGAAAGCAAAACATACAT GCATGCAAGG

Figure 4.11. Ectopic CENP-A^{Cnp1} incorporation occurs at regions enriched for asymmetric AT-rich sequences. (A) MEME (Bailey and Elkan 1994) identified a 40 bp asymmetric AT-rich motif enriched at 1 kb bins that show evidence of ectopic CENP-A^{Cnp1} incorporation. (B) This motif has an E-value of 2.4e-5 and is found in 17 of the 43 sequences used.

A



B

Name	Strand	Start	*p*-value	Sites
43	+	876	5.41e-14	ATGTGGGGGT GAGGGGGGAAAGAAAAATGCGGAGGAAG TGTTTCGGGA
42	+	90	2.27e-12	TACATAAAGG GAGGGGGGAATAAAAGAAACGAAGAAAG TAAAAGTAAA
28	+	354	5.71e-11	ATAACAAAAG GAGGGGAGAGAGTAATGAGAGACGCAAT ATTGTATAGA
22	+	757	1.68e-09	AGGTTAATTT GCAGGGAGAAAAAAAGGTTAGACGGAGG GCTGGAAGAG
27	-	709	1.87e-09	TGAGGGTAGC GAGGTCGAACAATGGGGAGCGGGGCAAA CTAGGGATTT
35	-	305	2.32e-09	GAGTAGCGAC GAGAAGAGAGAAAAAAGAGCGACAAAAG GAAATAAAAT
10	+	17	5.79e-09	AAGAATTGTA CAAGGAAGAAAGAAGAATTAGAGGCAAA ATATCTCAAT
37	+	859	9.33e-09	AAGGACTACC CCAGAAAGAGAGAGAGATATGAGGAAAA AAATTTTTGC
2	-	156	1.47e-08	AACTTGAGCC GAATAGGAAGAAAAGGAAACCATGAAAG TTCGGAAAGT
3	+	424	2.08e-08	CCCTTTAACT GAGGAAGGAATATCAGATAGCAGGCAAG AAGTAGTTAG
40	-	818	5.50e-08	GGATGCGTTG GAATGCGGAGTAGAGAAATTGAAGAGAG AGAGAGGACA
1	-	865	7.42e-08	CAATGCACAG GAGTATAGAGAAAGCAATGTGAATAAAT AGTATTATCA
21	+	557	8.59e-08	GTTGACAAAT GTGAAAGAAAAGAGAAAAGGGAATGAAT ATTTATAACA
23	+	632	1.14e-07	GCAATGGGGT GAGTGGGAAAAAAAGGGTTTGATCGGAT GCAAGTAGAC
13	-	310	1.14e-07	ATTTGACGCG AAAGGCCGACAAAGAAAAGTGGATAAAG TCAGCGCATA
33	+	241	1.50e-07	GCAGTGCGCA GTGGTAAAAGAGTGAAAATGGGGTGAAA TAACAAGGTG
31	-	789	2.83e-07	GAAGGAAATG GCGGGTAAACAGTAATAGAAGACGCAAC TTTGAAGGTT

Figure 4.12. Ectopic CENP-A^{Cnp1} incorporation occurs at regions enriched for asymmetric AT-rich sequences. (A) MEME (Bailey and Elkan 1994) identified a 28 bp asymmetric AT-and CG-rich motif enriched at 1 kb bins that show evidence of ectopic CENP-A^{Cnp1} incorporation. (B) This motif has an E-value of 9.3e-3 and is found in 17 of the 43 sequences used.

asymmetrical AT-rich sequences (Chuang and Kelly, 1999; Kong and DePamphilis, 2001; Lee et al., 2001). Furthermore, ORC has been detected within the central domain of fission yeast centromeres in the absence of any detectable replication origin activity (Hayashi et al., 2007; Segurado et al., 2003). Attempts to map the precise localisation of ORC within the central domain have not been possible because the original mapping data was low resolution and so far ChIP-seq has failed due to IP efficiency that was not sufficiently high. Therefore, an alternative approach was to investigate whether ectopic CENP-A^{Cnp1} incorporation occurs at or near replication origins.

A list of described replication origins in fission yeast were downloaded from OriDB (Siow et al., 2012). Origins with dubious support were excluded from this analysis, but were considered subsequently. Mean R_{IP} values were calculated for all likely and confirmed replication origins. R_{IP} values were compared to the chromosome wide distributions (Figure 4.13). Chromosome-wide distributions of R_{IP} values were distributed around a mean of ~1, indicating equal coverage in the CENP-A^{Cnp1} over-expression and the control samples. The long tail of high values results from sites where CENP-A^{Cnp1} is ectopically incorporated. Replication origins have a similar distribution of mean R_{IP} values with a mean value of ~1. There are a small number of origins that have mean R_{IP} values above the cut-off threshold for significant ectopic incorporation, but these are all located in subtelomeric regions or the centromeric outer-repeats. Dubious replication origins were analysed in the same way and similarly the only origins with evidence of significant ectopic CENP-A^{Cnp1} incorporation were in subtelomeric regions or the outer-repeats. However it may be necessary to also analyse ectopic CENP-A^{Cnp1} incorporation within replication origins to determine whether there is spatially restricted incorporation at smaller regions within origins. These analyses are currently being undertaken. It remains to be determined what features are associated with, and perhaps promote, ectopic CENP-A^{Cnp1} incorporation at chromosomal sites other than the centromere and subtelomeric regions. Ongoing analyses are underway to determine whether the regions of ectopic CENP-A^{Cnp1} incorporation are associated with other nucleosome-depleted regions, such as near transcriptional start sites.

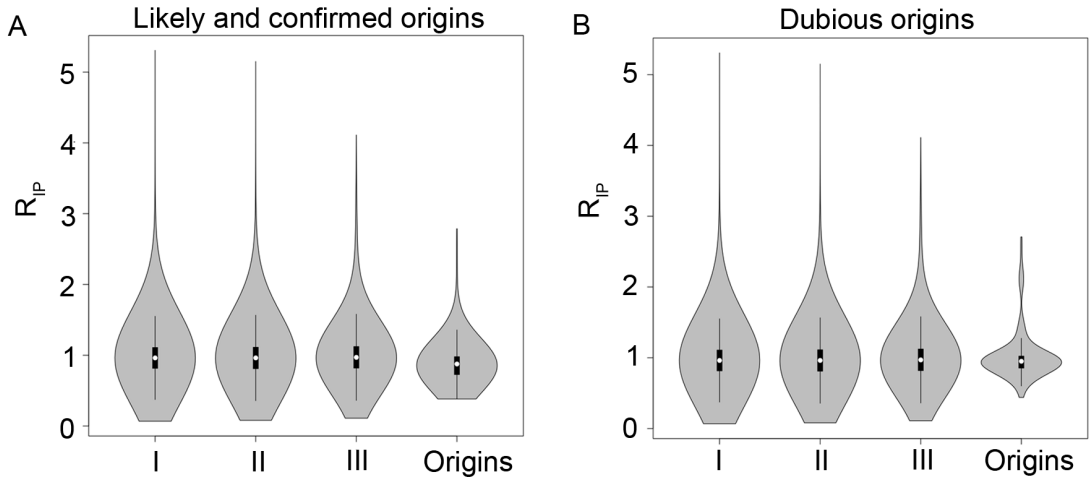


Figure 4.13. No ectopic CENP-A^{Cnp1} incorporation is detected at replication origins. (A) At likely and confirmed replication origins the mean R_{IP} values are distributed around a vaule of ~1, similar to the chromosome-wide distributions (on chromosomes I, II, and III). (B) Likewise, at dubious replication origins R_{IP} values are distributed around a vaule of ~1, similar to the chromosome-wide distributions (I, II, and III). In A and B the small number of origins with significant R_{IP} values consistent with ectopic CENP-A^{Cnp1} incorporation are found in the centromeric outer-repeats and subtelomeric regions.

4.2.9 No ectopic CENP-A^{Cnp1} incorporation is detected over genes

Broad domains of CENP-A^{Cnp1} enrichment following over-expression are limited to the subtelomeric regions and centromeres, but CENP-A^{Cnp1} incorporation might also occur on a more local scale. It remains possible that transcription may lead to H3 nucleosome eviction and their replacement by CENP-A^{Cnp1} containing nucleosomes. To test this possibility average R_{in} and R_{IP} values were calculated for annotated features in the *S. pombe* genome, i.e. for a single feature the average R_{IP} value was calculated (Figure 4.14). Features within all three centromere regions and within 100 kb of the telomeres were excluded to prevent incorporation at these sites from affecting the results.

R_{in} values for all feature types are distributed around a median that ranged between 0.981 for coding regions and 1.067 for repeat regions, which is very close to a value of 1 that indicates no difference between samples (Figure 4.14). Short nuclear RNAs had a median of 1.137, but only had a sample size of 7. Also, few individual features were found to be significantly different from the genome-wide mean using the p-value significance thresholds. For all genomic features the 25th and 75th quantiles were well below the p-value significance thresholds.

R_{ip} values showed a larger variation between different types of features. However, there is not evidence of widespread enrichment at any type of feature. The median value for coding regions (1.006), exons (0.965), genes (0.972), non-coding RNAs (0.960), and pseudogenes (1.061) were close to the genome-wide background with 25th and 75th quantiles that were well within the p-value significance thresholds. The remaining features have lower R_{ip} values, with repeat regions, snoRNAs, and tRNAs having 25th percentiles slightly below the lower p-value significance threshold. As this corresponds to CENP-A^{Cnp1} incorporation following over-expression that is lower than background levels it is unclear what relevance this has.

Ectopic CENP-A^{Cnp1} incorporation at a subset of genes was previously observed following CENP-A^{Cnp1} over-expression when a component of the FACT complex was mutated using microarrays to generate a genome-wide profile (Choi et al., 2012). Over-expression of CENP-A^{Cnp1} in wild-type cells was not found to result in ectopic incorporation at genes (Choi et al., 2012). Thus the results obtained here using ChIP-seq agree with these previous observations obtained by ChIP-chip. It would be

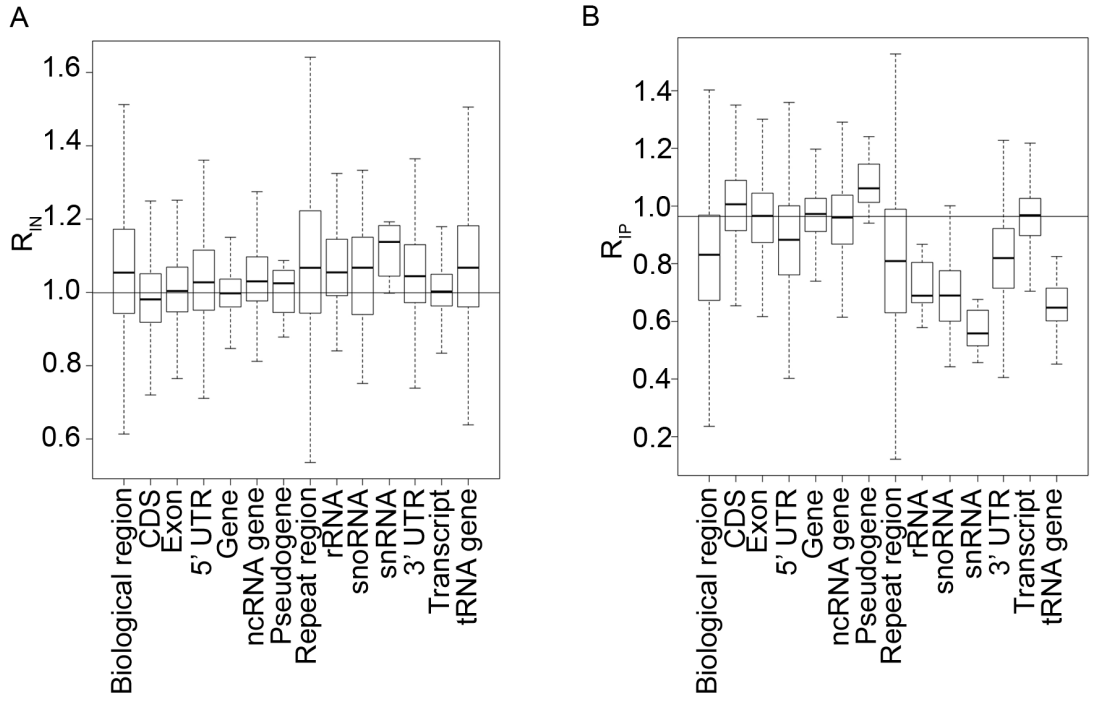


Figure 4.14. There is no ectopic CENP-A^{Cnp1} incorporation associated with annotated features. (A) R_{IN} values are closely distributed around a value of 1 for all annotated regions in the genome. This indicates that strains with CENP-A^{Cnp1} over-expressed and strains with wild-type CENP-A^{Cnp1} expression have similar extraction and sequencing biases for the sequenced IN. (B) R_{IP} values for most groups of annotated features were still distributed around a value of 1, but some features had lower values. No features showed evidence of significant ectopic CENP-A^{Cnp1} incorporation.

possible to analyse the relationship between gene expression level and ectopic CENP-A^{Cnp1} incorporation to determine if there is a subset of genes that show incorporation. This analysis is being undertaken.

4.3 Discussion

Ectopic incorporation of CENP-A^{Cnp1} following over-expression provides insight into the factors influencing CENP-A^{Cnp1} incorporation more generally. Over-expression of CENP-A leads to widespread incorporation across the genome in mammalian cells (Lacoste et al., 2014). In contrast, as shown here and indicated by previous studies (Castillo et al., 2013; Choi et al., 2012), in fission yeast ectopic incorporation following over-expression of CENP-A^{Cnp1} is primarily limited to the pericentromeric heterochromatin and subtelomeric regions with no significant incorporation at genes. Reads cannot be mapped uniquely to individual pericentromeric repeats so it is not known exactly how CENP-A^{Cnp1} and heterochromatin are organised over these regions. CENP-A^{Cnp1} may spread beyond the tRNA boundaries at each edge of the central domain and displace heterochromatin from nearby repeat elements or there may be interspersed regions of CENP-A^{Cnp1} and heterochromatin or cell to cell variation in the presence or absence of CENP-A^{Cnp1} on flanking outer repeats.

Limited significant incorporation was also detected at other sites in the genome but this ectopic incorporation was not found to be associated with any annotated genomic features. CENP-A^{Cnp1} incorporation at these locations is likely to be directed by chromatin features, or nuclear organization of chromosomal domains, since most ectopic incorporation is associated with the centromeric outer-repeats and subtelomeric regions rather than with any particular sequences. However, potential sequence features that may play a supplementary role were also identified, with asymmetric AT-rich sequences enriched at other regions of significant ectopic incorporation. It has previously been shown that telomere repeat sequences and associated heterochromatin can recruit ectopic CENP-A^{Cnp1} at an internal locus (Castillo et al., 2013). It should be noted that additional CENP-A^{Cnp1} may also be incorporated within the central domain of centromeres in over-expressing cells, but these cannot be detected using this methodology applied. Thus the amount of ectopic incorporation is a conservative estimate. Calculating actual increased enrichment levels across the central domain by ChIP-seq would require the use of multiple known concentrations of spiked-in CENP-A^{Cnp1} chromatin assembled on

DNA that is not present in *S. pombe*. This would allow ChIP-seq coverage to be related directly with the amount of DNA present so that the coverage could then be normalised to reflect the actual level of enrichment.

There was a slight shift in CENP-A^{Cnp1} occupancy within the central domain towards sites that are predicted to be occupied by nucleosomes. There are a number of explanations for this shift towards higher coverage on these predicted nucleosomal sites. Higher occupancy by CENP-A^{Cnp1} nucleosomes may lead to a slightly higher enrichment with reduced noise, thus allowing CENP-A^{Cnp1} nucleosomes to be mapped with higher resolution. Alternatively, since higher coverage was also obtained at existing CENP-A^{Cnp1} peaks, higher occupancy by CENP-A^{Cnp1} nucleosomes may lead to greater positional constraints at sites flanking highly positioned nucleosomes and result in statistical positioning of nearby nucleosomes. These observations and possibilities remain to be verified at higher resolution and tested. Until then, all explanations are speculative.

The most recent estimates of CENP-A^{Cnp1} occupancy within the central domain suggested that CENP-A^{Cnp1} occupies approximately 50% of potential sites in any given cell (Lando et al., 2012). This number does not include sites that are predicted to be occupied by nucleosomes, but where CENP-A^{Cnp1} occupancy has not been observed *in vivo*. The potential increase in CENP-A^{Cnp1} levels on the central domain in cells overexpressing CENP-A^{Cnp1} detected by ChIP-qPCR indicated that CENP-A^{Cnp1} may occupy additional sites within the central domains. The shift in occupancy detected suggests that CENP-A^{Cnp1} may also be occupying predicted sites that are not normally observed to be occupied *in vivo* in cells expressing wild-type CENP-A^{Cnp1} levels. Therefore, CENP-A^{Cnp1} may be occupying additional sites that are normally unoccupied (or occupied by other proteins or a protein complex) in wild-type cells. At the resolution utilised additional CENP-A^{Cnp1} occupied peaks could not be directly observed, but the shifts in occupancy within the central domain support such a hypothesis.

Subtelomeric regions are also the known sites of neocentromere formation (Ishii et al., 2008; Ogiyama et al., 2013). It was therefore of particular interest to determine whether ectopic CENP-A^{Cnp1} incorporation occurred within subtelomeric regions and how this may relate to neocentromere formation. Ectopic CENP-A^{Cnp1} incorporation

was observed within the subtelomeric regions of chromosome I, overlapping with the region where neocentromeres form. Increased ectopic CENP-A^{Cnp1} incorporation at these locations indicates that these may be sites of preferential incorporation that can seed neocentromere formation following canonical centromere deletion. On chromosome I this ectopic incorporation occurred over domains that extended towards the telomeres, while neocentromere formation occurs only over a segment of this ectopic incorporation domain that does not extend to the telomeres. This observation suggests that there may be mechanisms that limit the region continually occupied by CENP-A^{Cnp1} at stable neocentromeres. Transcription of genes has been shown to interfere with and thus shift neocentromere location in *Candida albicans* (Thakur and Sanyal, 2013). Thus it is possible that the transcription of flanking genes acts as a boundary constraining neocentromere location in these subtelomeric regions in *S. pombe.*

It has previously been observed that neocentromeres do not form near the telomeres of chromosome III, probably due to the presence of rDNA repeats (Ogiyama et al., 2013). Instead, unstable neocentromeres form at internal sites, which are resolved by chromosome fusion or rearrangements or rDNA repeat reduction (Ogiyama et al., 2013). Ectopic CENP-A^{Cnp1} incorporation was observed at subtelomeric loci on chromosome III and also on the internal chromosomal region known to form a neocentromere (cd385) adjacent to the rDNA repeat array. However, only limited ectopic CENP-A^{Cnp1} incorporation was observed at the internal site corresponding to that where the unstable cd385 neocentromere is known to form. It is possible that this small region of ectopic CENP-A^{Cnp1} incorporation identifies a site that acts as a seed for weak neocentromere formation in this region. Such a seed might trigger additional CENP-A^{Cnp1} chromatin assembly leading to neocentromere formation upon deletion of the chromosome III centromere. However, the low level of CENP-A^{Cnp1} enrichment detected is relatively common across other regions of the genome that do not form neocentromere. It is possible that many internal sites are capable of neocentromere formation on chromosome III and that this particular region was recovered as a functional neocentromere by chance. It is possible that other sites of ectopic CENP-A^{Cnp1} incorporation might also correlate with loci that are amenable to neocentromere formation.

Chapter 5: A high-resolution sequence-based map of CENP-A^{Cnp1}/kinetochore interactions

5.1 Introduction

Most regional centromeres form on repetitive sequences, while some evolutionarily young neocentromeres form on non-repetitive sequences (Fukagawa and Earnshaw, 2014). Therefore, positional information cannot be mapped to a unique location within most centromeres. In contrast, fission yeast (*S. pombe*) centromeres are comprised of a non-repetitive central core flanked by inner- and outer-repeat elements (Clarke and Baum, 1990; Clarke et al., 1986; Fishel et al., 1988; Hahnenberger et al., 1991; Murakami et al., 1991; Nakaseko et al., 1987). CENP-A^{Cnp1} occupies the central domain so its position can be mapped relative to underlying sequences (Partridge et al., 2000; Takahashi et al., 2000). CENP-A^{Cnp1} nucleosomes in fission yeast are highly positioned based on sequence features, so interactions with kinetochore components and other centromeric proteins can be mapped relative the CENP-A^{Cnp1} and the underlying sequence.

In budding yeast a single CENP-A^{Cse4} nucleosome interacts with a single spindle microtubule at each centromere (Winey et al., 1995). In contrast, in fission yeast and mammals multiple microtubules associate with each centromere (Ding et al., 1993; Dong et al., 2007; McEwen et al., 1997, 2001; Wendell et al., 1993; Zinkowski et al., 1991). Multiple kinetochore microtubules may interact with a single centromere via repeat subunits where each microtubule interacts with a functional kinetochore subunit (Zinkowski et al., 1991) or alternatively the kinetochore may form a network lacking a defined subunit structure (Dong et al., 2007). ChIP-seq analysis of fission yeast CENP-A^{Cnp1} indicates that there are ~20 sites occupied by CENP-A^{Cnp1} nucleosomes per centromere in a cell population. However single cell analysis using high-resolution microscopy estimated that the number of CENP-A^{Cnp1} nucleosomes present per centromere was only sufficient to occupy ~50% of sites in any given cell, but this estimate was subject to a large amount of uncertainty on the exact number of CENP-A^{Cnp1} nucleosome present (Lando et al., 2012). Thus, CENP-A^{Cnp1}

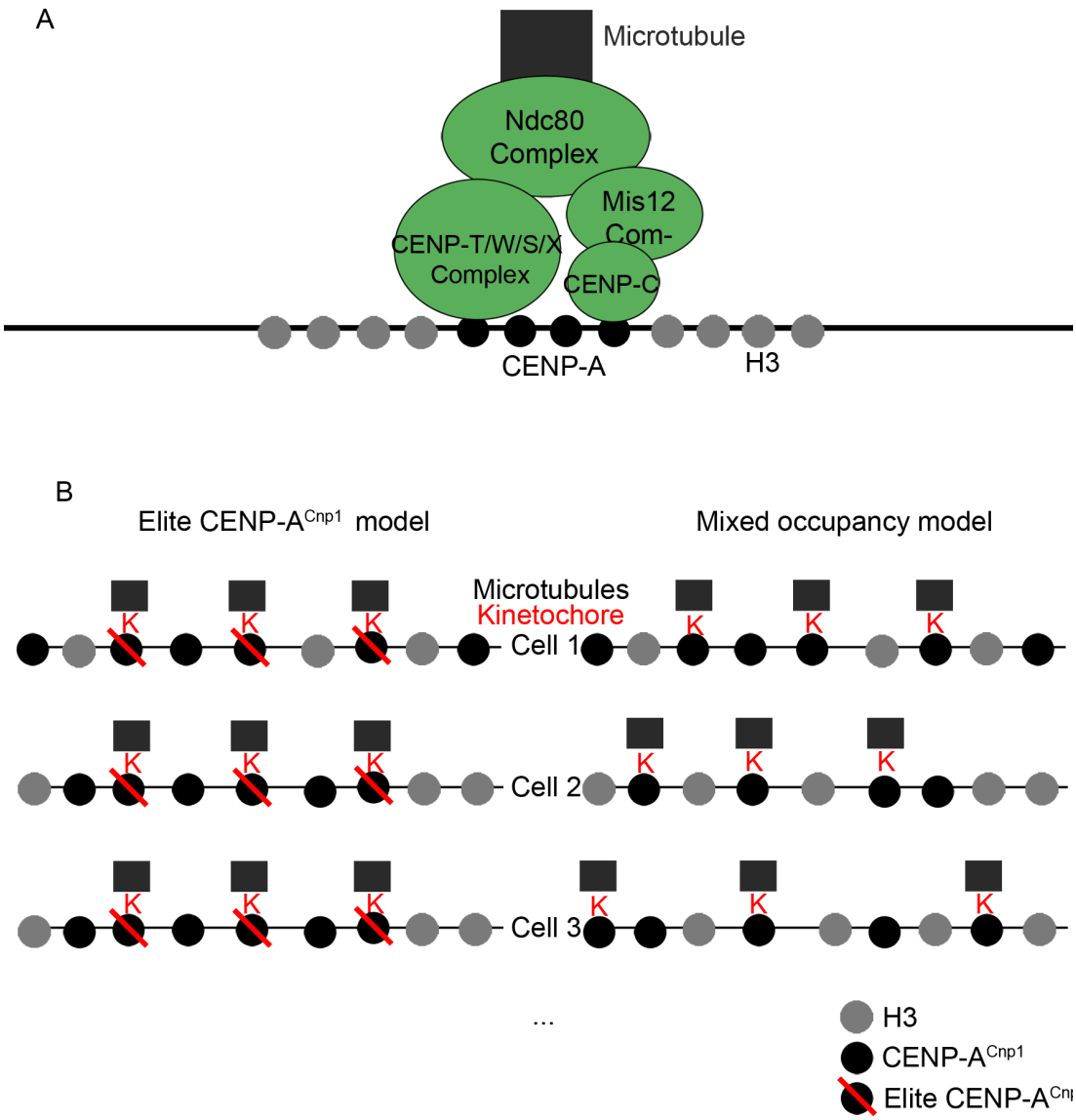


Figure 5.1. Structure of the kinetochore. (A) A general diagram of the kinetochore structure is shown. Association with the microtubule is mediated by the Ndc80 complex. There are two pathways bridging the CENP-A chromatin and the Ndc80 complex. These interactions are via the CENP-T/W/S/X complex or CENP-C and the Mis12 complex. (B) There are different ways that fission yeast CENP-A^{Cnp1} nucleosomes could interact with the kinetochore. In an elite CENP-A^{Cnp1} interactions with the kinetochore occur at a subset of CENP-A^{Cnp1} nucleosomes in every cell, despite additional CENP-A^{Cnp1} nucleosomes being present at other sites. In a mixed occupancy model the kinetochore interacts with different CENP-A^{Cnp1} nucleosomes in different cells. The kinetochore itself may form a single ordered structure that interacts with multiple microtubules. H3 nucleosomes are shown occupying sites not occupied by CENP-A^{Cnp1} nucleosomes, but the prevalence of H3 nucleosomes within the central domain is currently unknown.

nucleosomes are present at fission yeast centromeres in excess relative to the number of associated kinetochore microtubules, although perhaps not at an adequate number to fully occupy potential sites. It is possible that a subset of CENP-A^{Cnp1} nucleosome positions always interact with the kinetochore as "elite" CENP-A^{Cnp1} nucleosomes or alternately that all CENP-A^{Cnp1} nucleosome positions could potentially interact with the kinetochore in different cells in a mixed population in a stochastic manner (Figure 5.1B).

It is also of interest to determine if a protein or complex binds to the DNA between CENP-A^{Cnp1} nucleosomes. The vertebrate CENP-T/W/S/X complex has been proposed to bind DNA directly in a nucleosome-like structure (Nishino et al., 2012). This makes the fission yeast CENP-T^{Cnp20}/W/S/X complex a potential candidate for a complex associated with the gaps between CENP-A^{Cnp1} nucleosomes. Another candidate for a protein that might associate with the gaps detected between CENP-A^{Cnp1} nucleosomes is the CENP-A^{Cnp1} chaperone Scm3, which is homologous to the human CENP-A chaperone HJURP (Barnhart et al., 2011; Foltz et al., 2009; Hayashi et al., 2004; Pidoux et al., 2009; Sanchez-Pulido et al., 2009; Shuaib et al., 2010). *S. cerevisiae* Scm3 protein contains a DNA binding domain that binds non-specifically to AT-rich sequences and a predicted AT-hook domain, but the predicted AT-hook domain is absent from the *S. pombe* Scm3 protein (Aravind et al., 2007; Xiao et al., 2011).

Mis16 and Mis18 are required for CENP-A^{Cnp1} localisation to the centromere (Hayashi et al., 2004) and human Mis18α/β and Mis18BP1 also acts as loading factors for CENP-A (Fujita et al., 2007). No homolog of Mis18BP1 has been identified in *S. pombe*, but the novel protein Eic1 has been proposed to perform the function of recruiting the Mis16/Mis18 complex to centromeres (Subramanian et al., 2014). It was of interest to determine whether the localisation pattern of Eic1 was consistent with this function.

Mapping the localisation of kinetochore components relative to the underlying centromere sequence should allow several questions to be addressed. The study presented in this chapter sought to determine whether the CENP-T^{Cnp20}/W/S/X complex associates with the DNA residing between CENP-A^{Cnp1} nucleosomes and may therefore bind to DNA as has been proposed for the vertebrate complex. More

generally, it was investigated whether other centromere or kinetochore proteins are preferentially associated with these gaps. In addition, the analyses performed sought to determine if kinetochore proteins interact equivalently with all CENP-A^{Cnp1} nucleosomes, or whether there is a subset of elite CENP-A^{Cnp1} nucleosomes that interact with the kinetochore and evaluated the proposed role of Eic1 as a CENP-A^{Cnp1} recruitment factor.

5.2 Results

5.2.1. Enrichment of kinetochore proteins can be detected at the centromere

Unlike most regional centromeres studied the fission yeast centromere is an evolutionarily ancient centromere that forms on non-repetitive sequences. This allows positional information to be mapped specifically to a unique location within centromeres. It is known that patterns of CENP-A^{Cnp1} occupancy correlate with known kinetochore components since they all map to the central domain within each centromere (Ishii et al., 2008; Ogiyama et al., 2013; Partridge et al., 2000; Takahashi et al., 2000). However, the spatial relationship between these proteins relative to the underlying sequence has not been examined at high-resolution.

Strains of fission yeast expressing proteins of interest that had previously been tagged with GFP and expressed from their native promoter were subject to ChIP to study their localisation relative to each other and to the underlying DNA sequence. To determine whether all CENP-A^{Cnp1} nucleosomes or only an elite subset of CENP-A^{Cnp1} nucleosomes interact with the kinetochore ChIP was performed with the outer-kinetochore component Ndc80-GFP, the Mis12 complex components Mis12-GFP and Mis14-GFP, and the inner-kinetochore components CENP-C^{Cnp3}-GFP and CENP-T^{Cnp20}-GFP. Additionally, to determine if the CENP-T^{Cnp20}/W/S/X complex associates with regions between CENP-A^{Cnp1} nucleosomes ChIP of CENP-T^{Cnp20}-GFP was performed. ChIP of Scm3-GFP was also performed to determine if it associated with the AT-rich regions between CENP-A^{Cnp1} nucleosomes. Finally, ChIP was performed on Eic1-GFP, Eic2-GFP, Mis16-GFP, and Mis18-GFP to evaluate whether Eic1 and Eic2 localisation was consistent with their proposed roles as Mis16/Mis18 recruitment or stabilising factors (Hayashi et al., 2014; Subramanian et al., 2014). Enrichment of each protein was initially tested by qRT-PCR at a centromeric site of known high CENP-A^{Cnp1} enrichment within the central core of the chromosome II centromere relative to a negative control locus (actin). All proteins

were highly enriched in two biological replications at this site relative to the actin gene (*act1*$^{+}$), although the level of enrichment differed by several orders of magnitude for different GFP tagged proteins (Figure 5.2). Mis14-GFP and Mis16-GFP had the lowest level of enrichment (~15 fold enriched relative to *act1*$^{+}$), while CENP-T^{Cnp20} had the highest detectable enrichment level (>400 fold enriched relative to *act1*$^{+}$). Importantly, a mock IP control done using Dynabeads and anti-GFP antibody in a wild-type strain that did not contain any GFP-tagged proteins showed no enrichment of the centromere sequence relative to *act1*$^{+}$. GFP tagged proteins were used in this experiment due to their availability and at the time the confounding influence of tag size on CENP-A^{Cnp1} occupancy was not known.

Ndc80 is an outer-kinetochore component, so the enrichment of centromere DNA in Ndc80-GFP ChIP must result from protein-protein cross-links with kinetochore components that are in closer proximity to DNA. The greater distance of Ndc80 from chromatin perhaps explains the lower enrichment of centromere DNA (Takeuchi and Fukagawa, 2012). Mis16 also associates with the histone acetyltransferase Hat1 (Tong et al., 2012) and has been observed to be enriched generally in chromatin throughout the cell cycle (Hayashi et al., 2004), which may explain its relatively low centromeric enrichment. Mis16, Mis18, and Scm3 disassociate from the centromere during mitosis (Hayashi et al., 2004; Pidoux et al., 2009). CENP- T^{Cnp20} is an inner-kinetochore component that is thought to directly bind DNA or interact with CENP-A or H3 nucleosomes in vertebrates, which may explain its relatively high enrichment (Foltz et al., 2006a; Nishino et al., 2012; Ribeiro et al., 2010). Similarly, CENP-C^{Cnp3} and Scm3 ChIP were highly enriched for central domain DNA in agreement with their direct interaction with CENP-A^{Cnp1} (Ando et al., 2002; Carroll et al., 2010; Hayashi et al., 2004; Pidoux et al., 2009; Sanchez-Pulido et al., 2009).

5.2.2. Sequencing data is high quality and unbiased

Centromere DNA was adequately enriched in all GFP-tagged protein ChIP samples and thus ChIP samples and input (IN) DNA samples were processed for high throughput sequencing. The resulting DNA was paired-end sequenced to produce 100 bp reads on an Illumina HiSeq2000 (Chapter 3). Paired end sequencing was performed to ensure that the spatial resolution would be adequate for precise mapping since the chromatin was sheared by sonication. ChIP of the GFP-tagged kinetochore components with anti-GFP used a slightly different protocol that yields

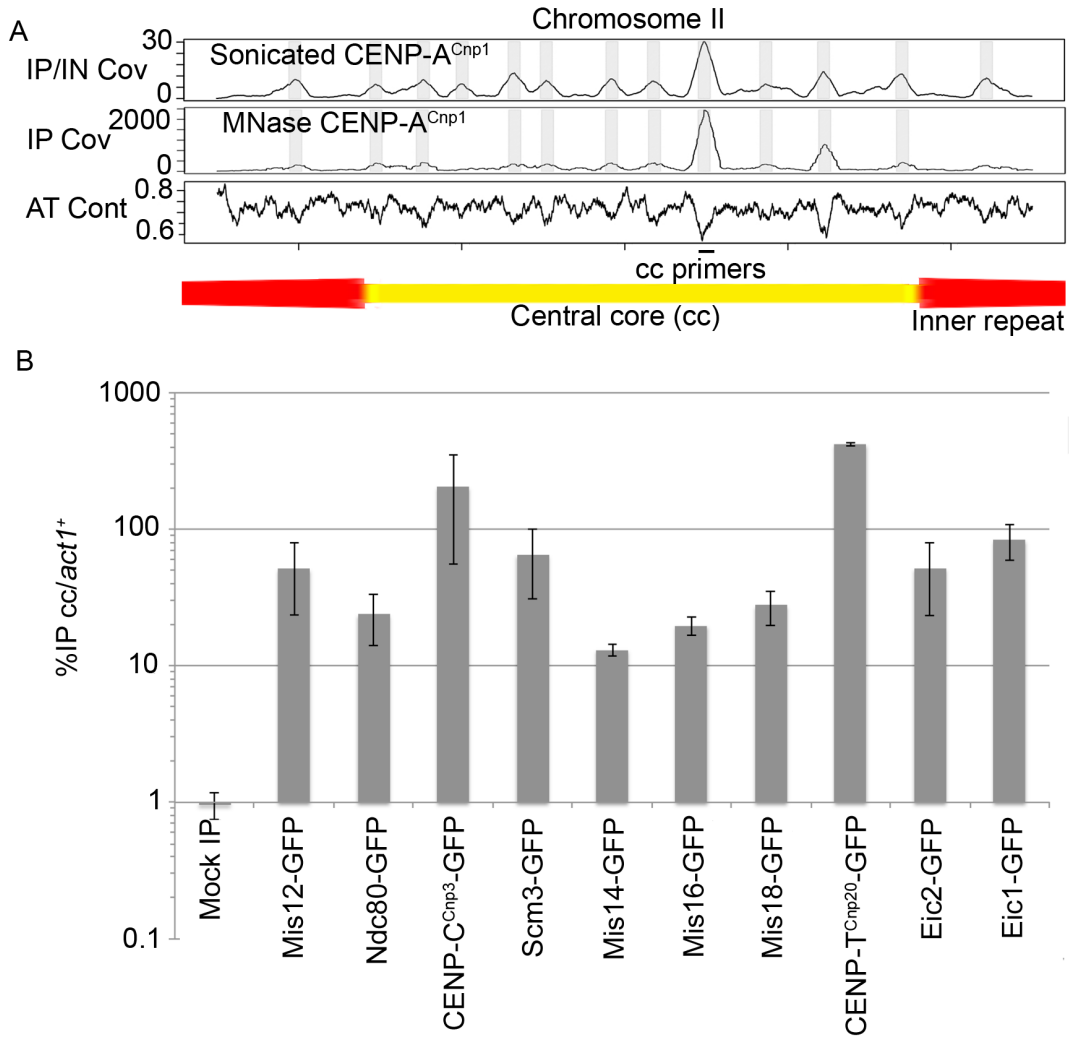


Figure 5.2. Kinetochore proteins are enriched at centromeres by qPCR. (A) Primers were chosen to amplify DNA from a CENP-A^{Cnp1} enriched peak region (Chapter 3) of the central core (cc) of the chromosome II centromere and a negative control locus, actin (*act1*$^{+}$). (B) GFP tagged kinetochore and centromere associated proteins were moderately to highly enriched at the central core relative to the negative control act1^{+} locus following ChIP, but a mock IP was not enriched (n=2). Error bars indicate standard error. The mock IP was done using Dynabeads and anti-GFP antibody in a wild-type strain that does not contain a GFP tagged protein.

higher enrichment and lower background than the protocol used for CENP-A^{Cnp1} ChIP. Unfortunately the same methodology cannot be applied to anti-CENP-A^{Cnp1} antiserum ChIP because it was found to be much less efficient. Given the known biases in Illumina sequencing, the quality of the sequenced data was first evaluated to determine whether the data was appropriate for accurate positional mapping of kinetochore proteins relative to CENP-A^{Cnp1} and the underlying sequence.

All GFP tagged protein ChIP-seq and IN data sets were of moderately high quality (CENP-T^{Cnp20}-GFP IN is shown in Figure 5.3 as a representative example). This included average Phred scores greater than 30 across the majority of sequenced reads, as well as Phred scores greater than 30 across the entire length of sequenced reads. Sequence content was unimodally distributed, with a mean AT-content consistent with the *S. pombe* genome, and uniform AT-content across the length of reads indicated that there was only minimal primer contamination with no degradation in quality along the length of reads.

It was of particular interest whether the data was unbiased and could accurately be used to map position at high resolution within the central domain, especially given the high AT-content of this region. Moreover, it had already been necessary to repeat CENP-A^{Cnp1} ChIP-seq experiments (Chapter 3) to obtain data that could be used for this purpose and that had the power to detect enrichment at all sites within centromeres. ChIP-seq to map the position of various kinetochore proteins must be capable of detecting enrichment over the AT-rich regions that had low CENP-A^{Cnp1} coverage to detect if any proteins contact these sites with high accuracy. All sequenced IN data was largely unbiased and showed no evidence of lower coverage at more AT-rich regions. For example, sequencing of the CENP-T^{Cnp20}-GFP strain IN material (Figure 5.4) shows that genome-wide coverage is unbiased relative to AT-content. This is observed as no correlation between IN coverage and AT-content at 10,000 random sites across the genome (τ=0.008; Figure 5.4A) and a negligible correlation at 10,000 random sites within the central domains (τ=-0.100; Figure 5.4B). Furthermore, there is no drop in coverage in the IN sample within the central domains relative to the rest of the genome (Figure 5.4C), so the previously detected extraction bias was not observed. It is not known what causes this difference, but it may be related to subtle changes in kinetochore interactions due to the presence of the GFP tag. Thus, the data sets generated were of high quality and

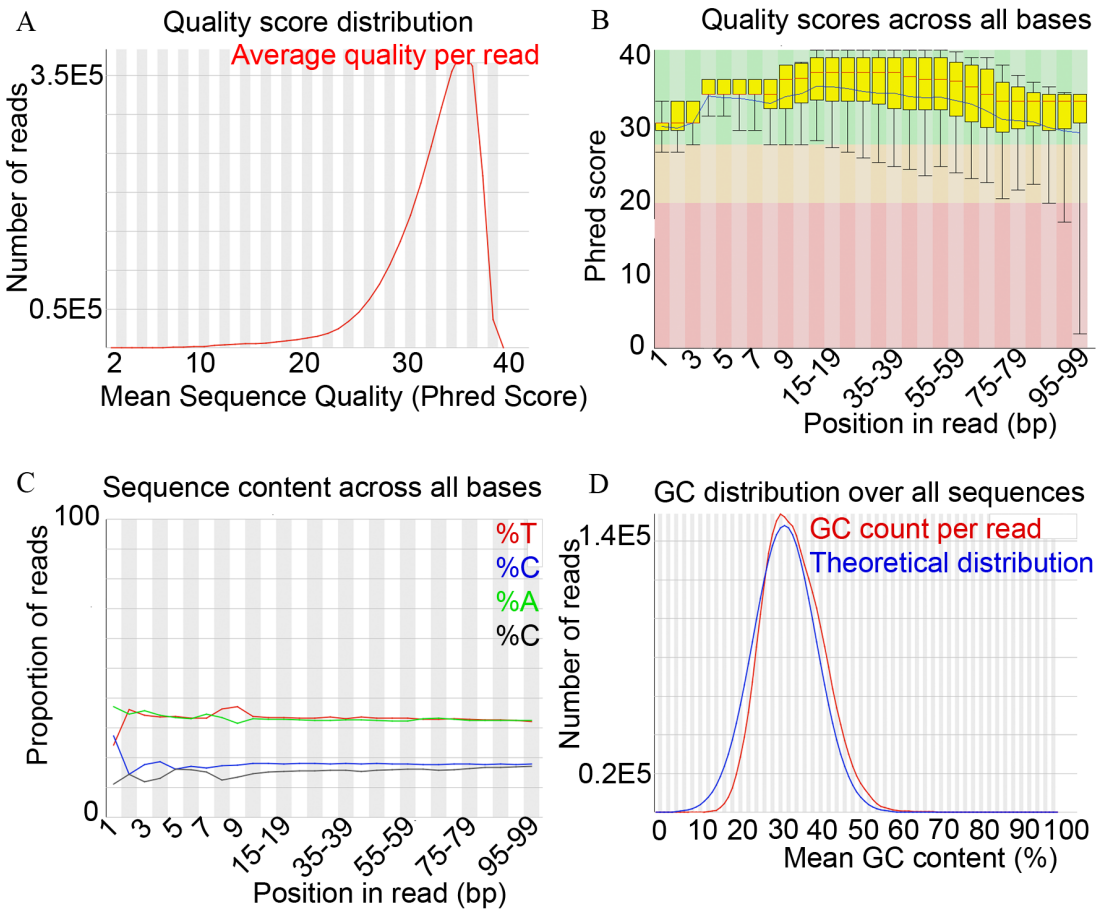


Figure 5.3. New Illumina sequencing data is high quality. Sequence quality information for illumina sequencing of CENP-T^{Cnp20}-GFP ChIP input is shown as a representative example. (A) Illumina sequencing resulted in reads with high quality, with most reads having an average Phred quality score >30. (B) Across the length of reads the quality remains high. (C) Nucleotide content is uniform across the length of reads (although there is evidence of minor primer contamination at the beginning of the reads) and (D) across reads the nucleotide content is unimodally distributed with values consistent with data from *S. pombe*.

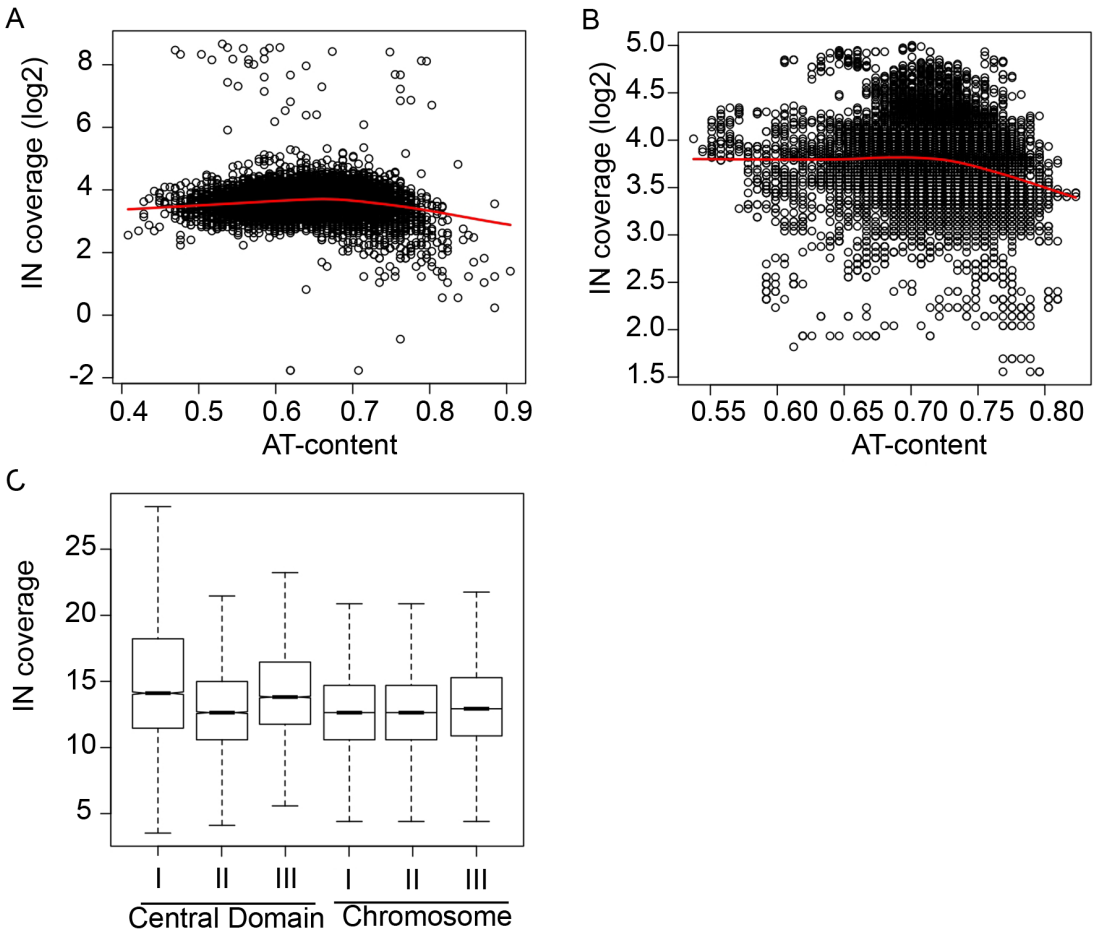


Figure 5.4. Illumina sequencing of input shows that sequenced reads are unbiased across the *S. pombe* genome. Sequencing and extraction bias is shown for illumina sequencing of the CENP-T^{Cnp20}-GFP ChIP input as a representative example. (A) At 10,000 random sites across the genome there is no correlation between illumina sequenced CENP-T^{Cnp20}-GFP IN coverage and AT-content (τ =0.008). (B) Similarly, at 10,000 random sites within the central domains the correlation between IN coverage and AT-content is negligible (τ=-0.100). Sequencing bias is thus not a concern and the data can be used to map ChIP-seq data within the central domains in an unbiased manner. (C) There is relatively uniform coverage across the central domains relative to the chromosome averages, so the data are not affected by a clear extraciton bias.

were deemed suitable for mapping the localisation of all GFP-tagged proteins across the centromeric central domains relative to both CENP-A^{Cnp1} and the underlying DNA sequence, at high resolution and in an unbiased manner.

5.2.3. Kinetochore components localise to discrete peaks that correlate with CENP-A^{Cnp1} nucleosomes

The ChIP-seq data was analysed to determine the localisation of the centromeric and kinetochore proteins relative to CENP-A^{Cnp1} and underlying sequence at high resolution. It was of particular interest to assess whether any of the kinetochore proteins mapped to the gaps between CENP-A^{Cnp1} occupied sites or to only a subset of CENP-A^{Cnp1} occupied sites.

ChIP-seq data of all the kinetochore components were moderately to highly enriched at the centromere relative to the rest of the genome (Figure 5.5). Fragment coverage across the centromere of chromosome II is shown in all figures since the centromeres of chromosome I and III share a region of homology and therefore reads cannot be mapped uniquely to these regions. Nevertheless, all analyses are done using data from all centromeres and the results for each centromere are consistent with other centromeres for all ChIP-seq samples, with the exception of Ndc80-GFP, which is analysed in more detail. This enrichment was specific to the central domains (Figure 5.6). GFP-tagged Scm3, Mis12, Mis18, Eic1, Eic2, CENP-C^{Cnp3}, and Ndc80 were enriched at specific discrete peaks within the central domains, while Mis14 and Mis16 showed low uniform enrichment across the central domains (Figure 5.7). ChIP-seq of H3 was also performed to determine how canonical nucleosomes are distributed over the central domain. However, because extremely coverage of centromere DNA resulted form this experiment the data was not analysed further (data not shown). The ChIP-seq data was normalised relative to the IN to control for sequencing and extraction biases. In most cases this did not affect the observed enrichment patterns, but Mis14-GFP and Mis16-GFP showed very narrow peaks of enrichment within the central domains that are caused by small local variations in the IP and IN coverage. However, Mis14-GFP and Mis16-GFP IP and IN coverage are both relatively uniform across the central domains. The peaks are an artefact of the normalisation procedure. There was a high correlation between enrichment of most of these kinetochore proteins with each other and with the distribution of CENP-A^{Cnp1} within the central domains (Figure 5.8). A relatively

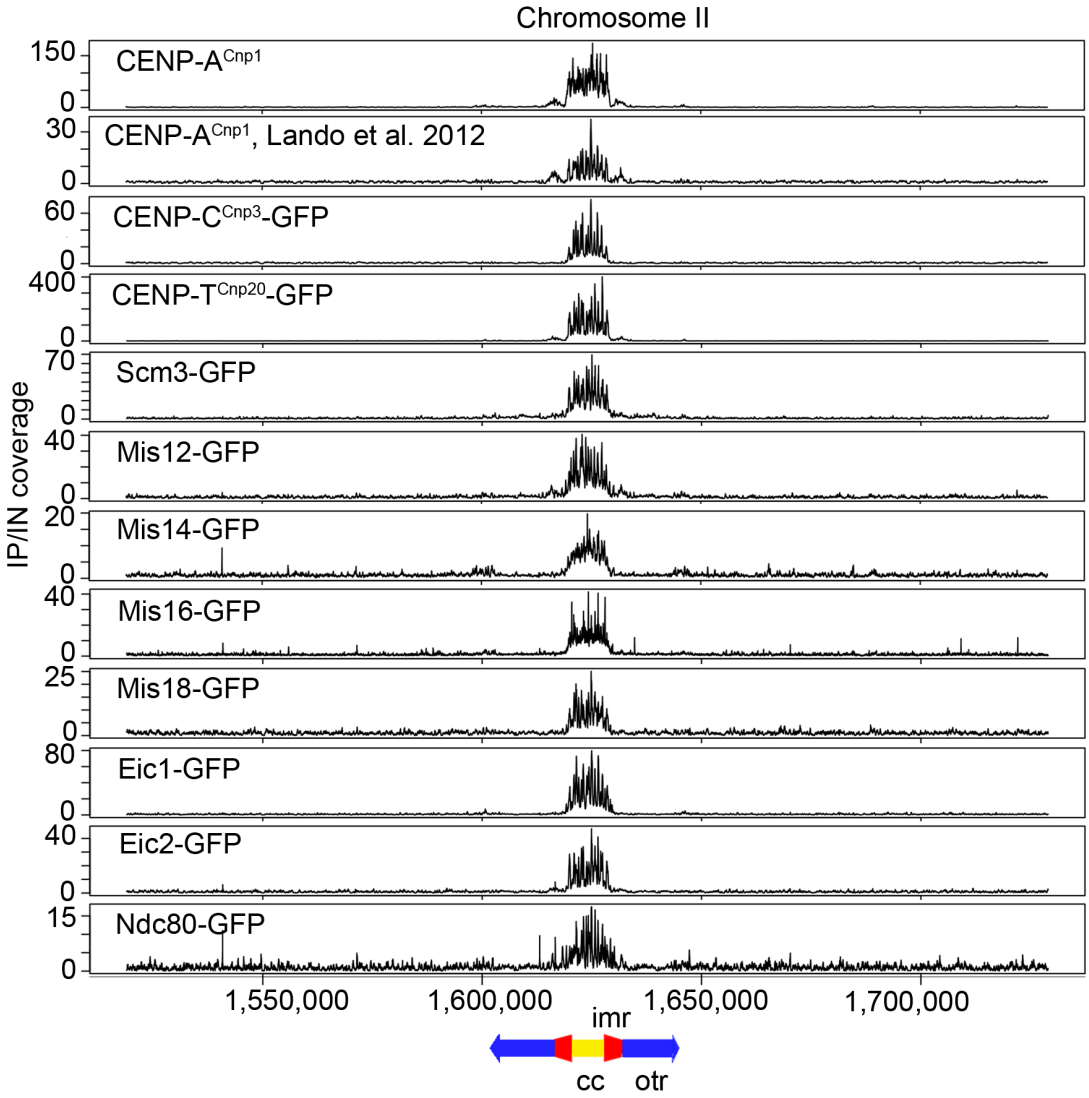


Figure 5.5. GFP tagged kinetochore and centromere proteins are enriched at the centromeres by ChIP-seq. IP/IN coverage is shown for ChIP-seq of GFP tagged kinetochore and centromeric proteins, including 100 kb of flanking sequences around the central domain of the chromosome II centromere as a representative example of centromeric enrichment. Additionally, published CENP-A^{Cnp1} ChIP-seq data (Lando et al. 2012) and new CENP-A^{Cnp1} ChIP-seq data (Chapter 3) are shown. All data shows moderate to high enrichment at centromeres.

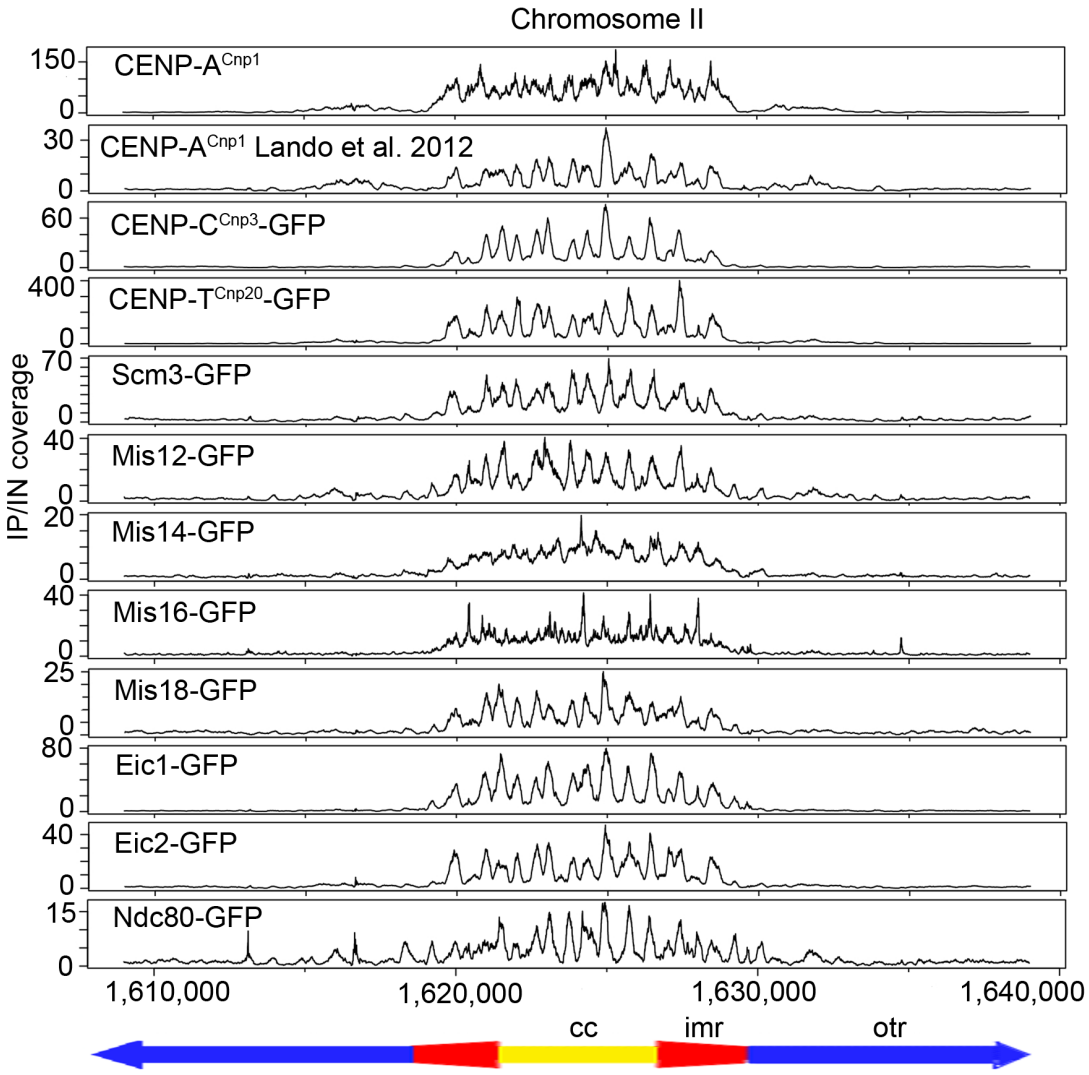


Figure 5.6. GFP tagged kinetochore and centromere associated proteins are highly enriched at the central domains. IP/IN coverage is shown for ChIP-seq of GFP tagged kinetochore and centromere associated proteins with 10 kb of flanking sequences around the central domain of the chromosome II centromere as a representative exmaple of centromeric enrichment. All data shows moderate to high enrichment specific to the central domains, the central core (cc) and part of the inner-repeats (imr). Ndc80 also shows some enrichment peaks extending beyond the central domains, but these may be due to noise from the lower level of enrichment.

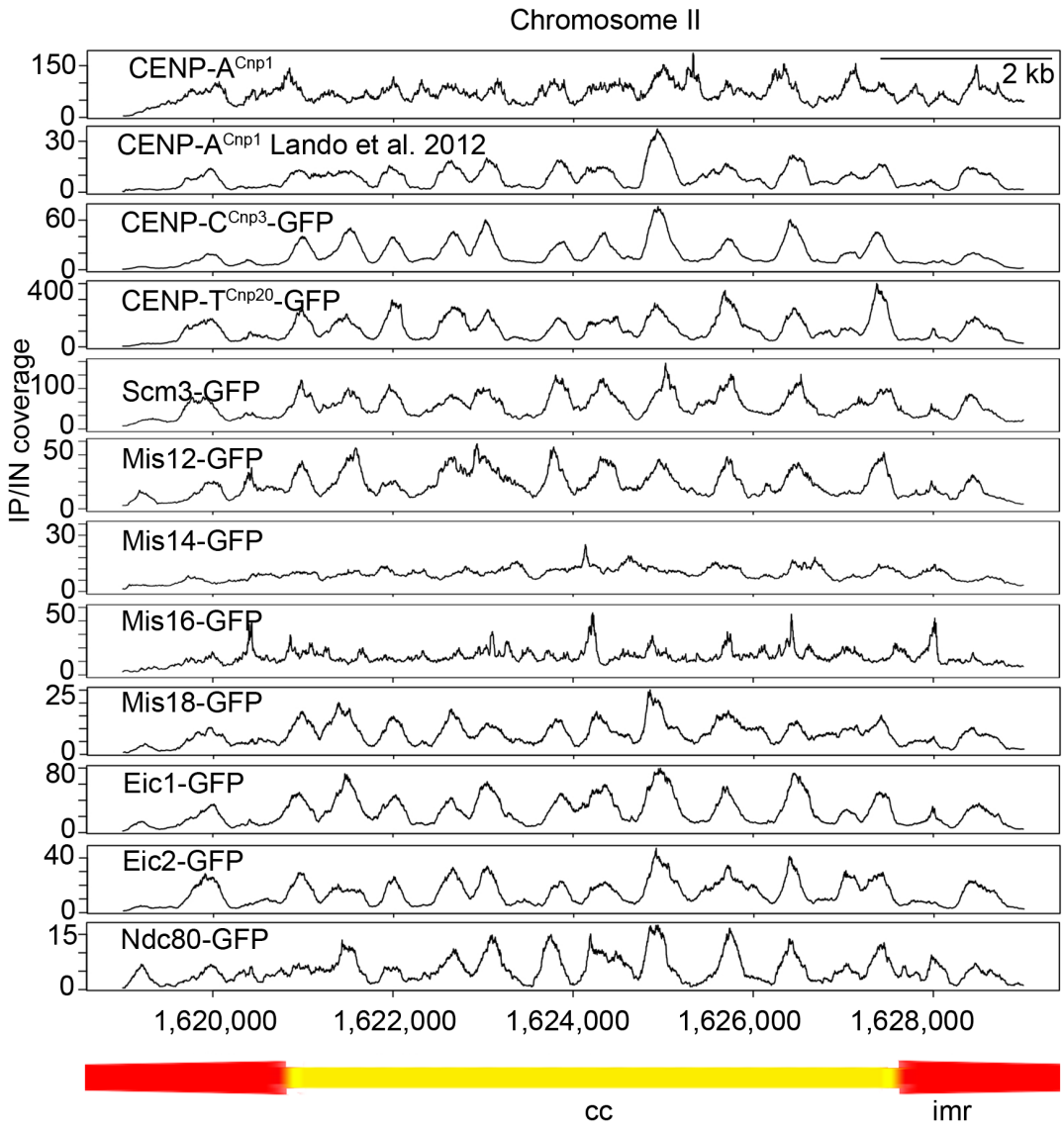


Figure 5.7. GFP tagged kinetochore and centromere associated proteins are enriched at discrete peaks within centromeres. IP/IN coverage is shown for ChIP-seq of GFP tagged kinetochore and centromere associated proteins with 0 kb of flanking sequences around the central domain of the chromosome II centromere as a representative example of centromeric enrichment. The data show distinct peaks of enrichment for the samples, with the exceptions of Mis14 and Mis16, which have a few artifactual peaks cause by small variations in the

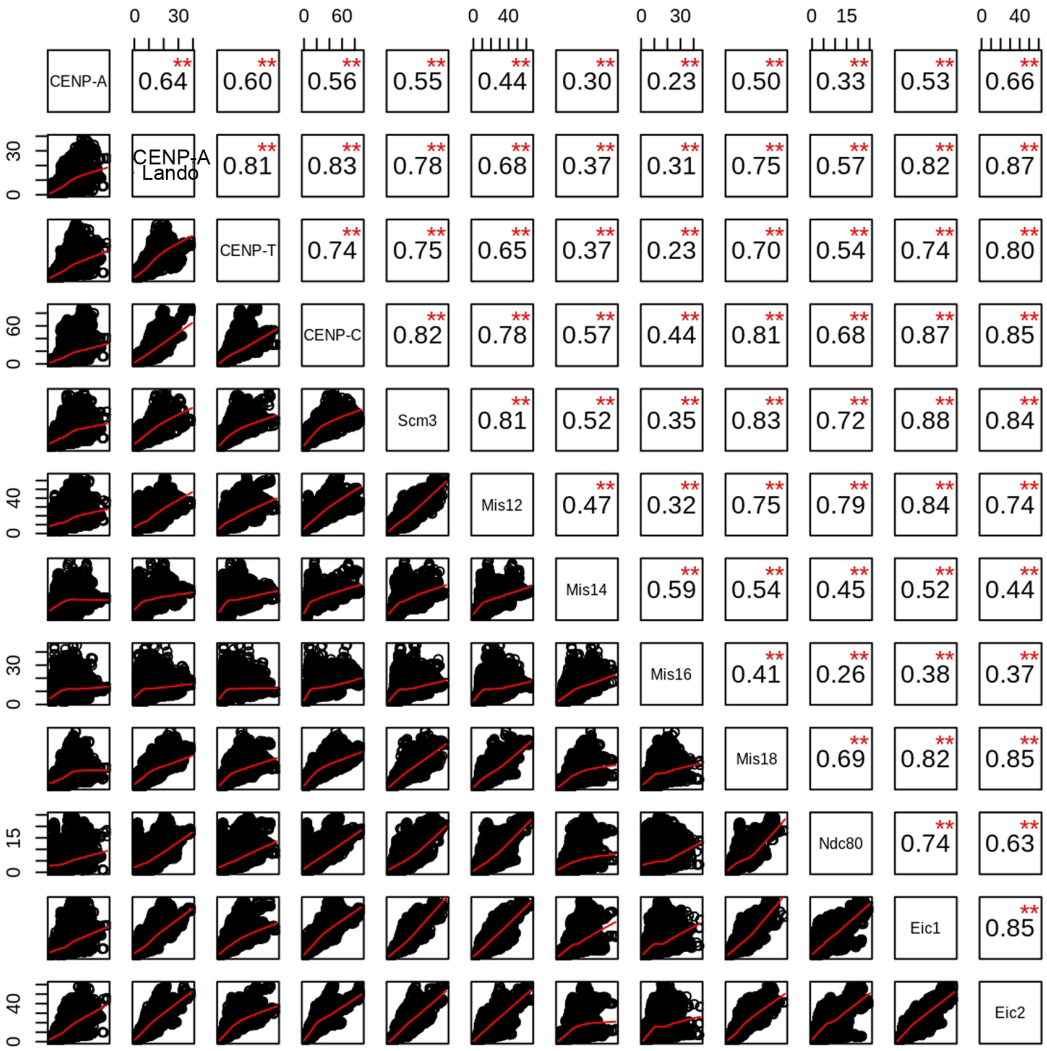


Figure 5.8. Centromere and kinetochore protein enrichment is correlated within the central domains. All pairwise correlations between IP/IN coverage of ChIP-seq data at the three central domains are shown. Most components are highly correlated with each other with a few exceptions. The newly generated CENP-A^{Cnp1} ChIP-seq data is only moderately correlated with other data due to the high level of enrichment over the entire central domains in addition to higher enrichment at specific peaks. Also Mis14 and Mis16 only show a low to moderate level of correlation with other data due to their relatively low levels of enrichment that are not associated with any peak enrichment. Asterisks represent p-values less than 0.001.

high correlation was observed between kinetochore protein distributions and the previously published sonicated CENP-A^{Cnp1} ChIP-seq association pattern (Lando et al., 2012). However, when compared with the newly generated CENP-A^{Cnp1} ChIP-seq data described in Chapter 3, only a moderate correlation between the distribution GFP-tagged kinetochore proteins and this new CENP-A^{Cnp1} ChIP-seq data was obtained. This difference is explained by the fact that in the new data set CENP-A^{Cnp1} was found to be highly enriched at all sites across the central domains of centromeres with only slightly higher enrichment at specific peaks. The Mis14-GFP and Mis16-GFP association patterns were only weakly correlated with those of the other kinetochore proteins analysed since both Mis14-GFP and Mis16-GFP only display a moderate level of enrichment over the central domains with no discrete peaks of enrichment.

The level of correlation varies between different components, but there were no components that were negatively correlated with CENP-A^{Cnp1} or other kinetochore components. Such a negative correlation would be expected for any protein that associated primarily with the gaps between CENP-A^{Cnp1} nucleosomes. Ndc80 is an outer-kinetochore component that interacts with spindle microtubules so enrichment of centromere DNA must result from protein-protein cross links with more DNA proximal kinetochore components. Association of any centromere or kinetochore component within the AT-rich gap regions would be expected to be reflected in an enrichment of these region in the ChIP of Ndc80-GFP or one of the kinetochore GFP-tagged components analysed here. However, since these regions were not enriched in the ChIP of any of the centromere-kinetochore proteins analysed it appears that none of the canonical kinetochore components are associated with these AT-rich regions.

Mis14-GFP and Mis16-GFP were the only components that showed an association pattern across the central domains that differed significantly from CENP-A^{Cnp1}. Both proteins were enriched broadly across the central domain with no clear peaks (Figure 5.7). This pattern is consistent with that expected for a gap-associated protein, but it remains far from clear whether these proteins associate with these gaps. Moreover, Mis12 and Mis14 are known to associate within the same complex and Mis16 and Mis18 are components of a distinct complex with Eic1 and Eic2 (Hayashi et al., 2014; Subramanian et al., 2014). It is not clear why these individual

components would have such different profiles if they are actually gap associated. Furthermore, Mis14-GFP and Mis16-GFP occupancies are not negatively correlated with CENP-A^{Cnp1} occupancy (Figure 5.8), as would be expected for any truly specific AT-rich gap associated protein, although extensive cross-linking to CENP-A^{Cnp1} could result in such a pattern. A more likely explanation is that the broad enrichment is due to low power since Mis14-GFP and Mis16-GFP have the lowest levels of enrichment as determined by qRT-PCR. Still, it remain inconclusive as to whether the less distinct coverage is due to low enrichment, or conversely if the low enrichment results from a different association profile. It would be possible to perform additional ChIP-Seq replicates to increase coverage, which might increase resolution if the pattern is due to lo enrichment. This is unlikely since both proteins belong to complexes that map to the same locations as CENP-A^{Cnp1}.

5.2.4 CENP-T^{Cnp20}/W/S/X does not preferentially associate with DNA between CENP-A nucleosomes and it remains unclear whether it contacts centromeric DNA directly

The large gaps between CENP-A^{Cnp1} peaks within the central domains of fission yeast centromeres suggest that another protein or complex may be more specifically associated with these regions. One possible candidate was the complex known as CENP-T/W/S/X in vertebrates, which has been proposed to form a DNA binding nucleosome-like complex, although the complex may interact with H3 or CENP-A nucleosomes (Foltz et al., 2006a; Nishino et al., 2012; Ribeiro et al., 2010). If DNA can directly wrap around a CENP-T/W/S/X in fission yeast nucleosome-like particle then it may be possible to visualise this difference via ChIP-seq if the complex interacts with different sequences or regions than CENP-A^{Cnp1} nucleosomes, such as localising to CENP-A^{Cnp1} gap regions. Following ChIP-seq CENP-T^{Cnp20} enrichment across centromere central domains was found to correlate with CENP-A^{Cnp1} occupancy. This CENP-T^{Cnp20} association pattern was analysed in more detail since it was a primary candidate for a component associated with the DNA sequences residing between CENP-A^{Cnp1} nucleosomes.

Previously published CENP-A^{Cnp1} ChIP-seq data (Lando et al., 2012) was used for the following analyses since it was more strongly correlated with coverage of kinetochore components within the central domains despite its higher bias. Also, predictions based on centromeric sequence features indicate that CENP-A^{Cnp1}

would be strongly positioned (Chapter 3), which agrees with the published data. CENP-T^{Cnp20} was found to localise to specific peaks within the central domains with large gaps between the peaks (Figure 5.9). This is similar to the localisation pattern of CENP-A^{Cnp1} peaks observed previously. A similar number of peaks are defined in both datasets, with totals of 54 CENP-A^{Cnp1} peaks and 48 CENP-T^{Cnp20} peaks at all three centromeres. The overlap between peaks was analysed by comparing the peaks of maximal enrichment and defining peaks as shared between the two data sets if they coincided within a limiting distance. Using this test significant overlap was detected between the peaks in the two datasets, even when the cut-off distance between peaks was small (Figure 5.10A). A range of values is shown for the maximum distance between shared peaks and how this affects the trade-off between sensitivity in detecting shared peaks and the specificity in ensuring the peaks occur at the same site. A maximum distance between peaks of 30 bp was chosen as conservative distance that identified most shared peaks. The two data sets share 33 peaks that co-occur within 30 bp of each other (Figure 5.10B). The small difference observed is due to a number of factors including variation in the location of maximum enrichment within a peak and imprecisely defined peak positions.

The IP/IN coverage ratios of CENP-A^{Cnp1} and CENP-T^{Cnp20} were strongly correlated at 10,000 random positions within the central domains of all three centromeres (τ=0.81; Figure 5.11). There is only a weak correlation, driven primarily by low sequence coverage regions, between CENP-A^{Cnp1} and CENP-T^{Cnp20} at 10,000 random sites across the genome (τ=0.41). The IN data were analysed for comparison to see whether extraction and sequencing biases could lead to a similar pattern. CENP-A^{Cnp1} and CENP-T^{Cnp20} IN coverage at 10,000 randomly chosen positions within the central domains of all three centromeres showed a weak correlation (τ=0.30). At 10,000 randomly selected locations from the entire genome there was also a weak correlation between CENP-A^{Cnp1} and CENP-T^{Cnp20} IN coverage (τ=0.32). This suggests that there is a slight bias that is shared between the two data sets, but the effect is likely minimal given the high levels of enrichment within centromeres. In particular, the CENP-T^{Cnp20} ChIP-seq data was largely unaffected by both sequencing and extraction biases, so the peak profile obtained is expected to be highly accurate. There was a strong positive correlation between enrichment and peak location for CENP-A^{Cnp1} and CENP-T^{Cnp20}. This is in striking

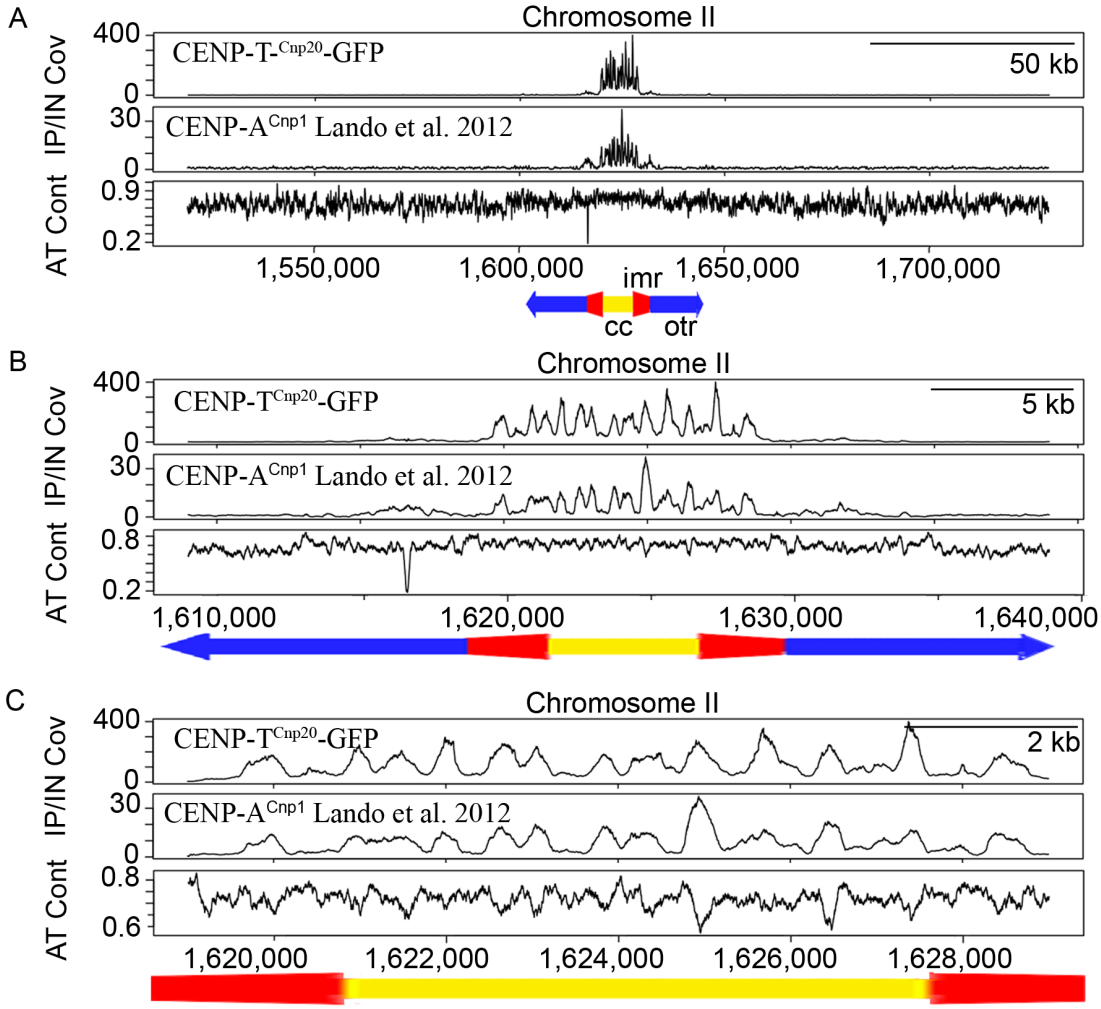


Figure 5.9. CENP-T^{Cnp20} and CENP-A^{Cnp1} are enriched at the same peaks within the central domains. CENP-A^{Cnp1} (Lando et al. 2012) and CENP-T^{Cnp20}-GFP IP/IN coverage are plotted at the chromosome II centromere, including (A) 100 kb, (B) 10 kb, and (C) 0 kb around the central domain, as a representative example of centomeric enrichment. Both localise specifically to discrete peaks within the central domains, composed of the central core (cc) and part of the inner-repeats (imr), and are absent from the outer repeats (otr).

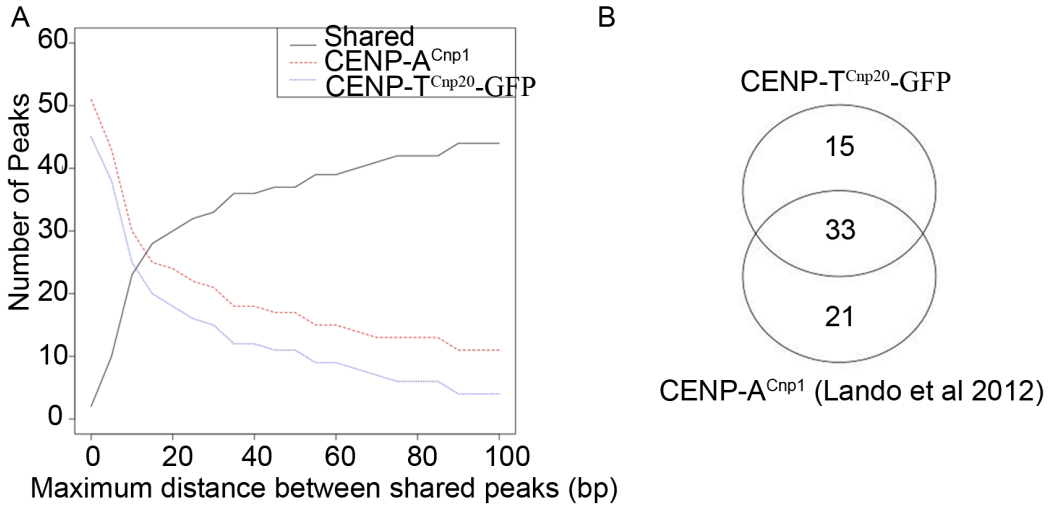


Figure 5.10. CENP-T^{Cnp20} and CENP-A^{Cnp1} are enriched at the same peaks within centromeres. (A) CENP-T^{Cnp20}-GFP and published CENP-A^{Cnp1} data (Lando et al. 2012) are enriched for peaks within all centromeres that are shared between data sets. Peaks are defined as shared if they are within a certain maximum distance of each other. A range of values for the maximum distance between shared peaks is shown. (B) 30 bp was used as the maximum distance between shared peaks as a trade-off between sensitivity in detecting shared peaks and specificity in constraining the distance between shared peaks. If shared peaks are defined as being within 30 bp of each other then 33 of the peaks overlap between the two data sets.

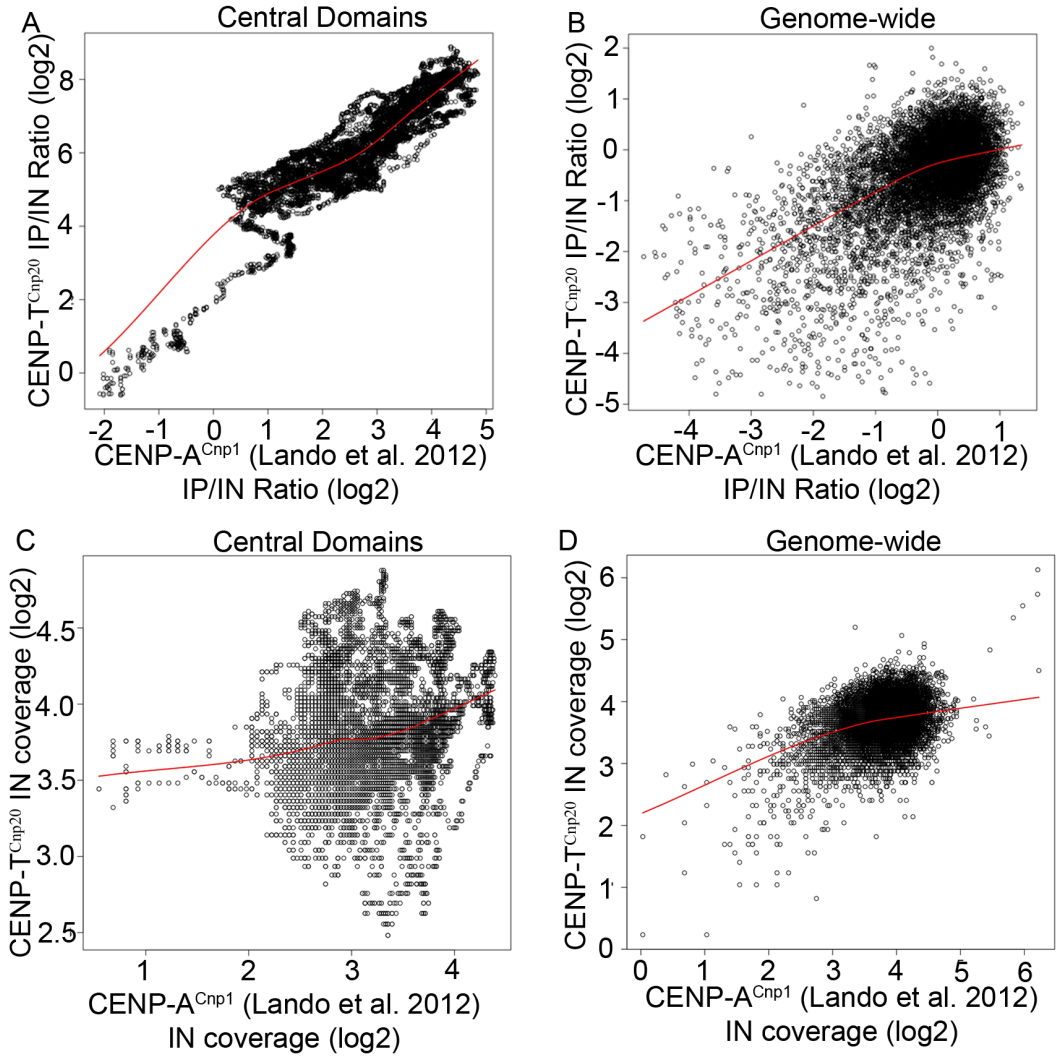


Figure 5.11. No evidence that CENP-T^{Cnp20} preferentially associates with DNA between CENP-A^{Cnp1} nucleosomes. (A) There is a strong correlation between the IP/IN ratio of CENP-T^{Cnp20} and the published CENP-A^{Cnp1} ChIP-seq data (Lando et al. 2012) at 10,000 random positions within the central domains (τ =0.81). (B) At 10,000 random sites genome-wide there is only a weak correlation between CENP-T^{Cnp20} and CENP-A^{Cnp1} coverage, which is driven primarily by sites with low coverage (τ=0.41). Sequencing bias is unlikely to explain the relationship between CENP-T^{Cnp20} and CENP-A^{Cnp1}. (C) There is a weak correlation in the IN coverage of CENP-T^{Cnp20} and the published CENP-A^{Cnp1} ChIP-seq data (Lando et al. 2012) at 10,000 random positions within the central domains (τ=0.30). (D) At 10,000 random sites genome-wide there is only a weak correlation between CENP-T^{Cnp20} and CENP-A^{Cnp1} IN coverage, which is driven primarily by sites with extremely low or high coverage (τ=0.41).

contrast to the expected negative correlation if CENP-T^{Cnp20} (and other components of the CENP-T/W/S/X complex) preferentially bound to DNA residing in the gaps between CENP-A^{Cnp1} nucleosomes.

Thus the data suggests that in fission yeast the CENP-T^{Cnp20} (and presumably other CENP-T^{Cnp20}/W/S/X components) does not preferentially associate with DNA residing between CENP-A^{Cnp1} nucleosomes. Thus, these analyses find no evidence that the CENP-T^{Cnp20}/W/S/X complex binds DNA, as it does not localise to the gaps in CENP-A^{Cnp1} occupancy. However, this does not exclude the possibility that the CENP-T^{Cnp20}/W/S/X complex binds DNA at the same sites as CENP-A^{Cnp1}. Therefore either the complex binds DNA by localising to the same sites as H3 or CENP-A^{Cnp1} nucleosomes or it localises through indirect interactions with CENP-A^{Cnp1} or H3.

5.2.5. Scm3 does not preferentially interacts with the AT-rich DNA between CENP-A^{Cnp1} nucleosomes

A second candidate for a protein localising to the gaps between CENP-A^{Cnp1} nucleosomes is the CENP-A^{Cnp1} chaperone Scm3. Scm3 and its human ortholog HJURP (Holliday junction recognizing protein) act as chaperones for CENP-A centromere localisation (Barnhart et al., 2011; Foltz et al., 2009; Hayashi et al., 2004; Pidoux et al., 2009; Sanchez-Pulido et al., 2009; Shuaib et al., 2010). *In vitro* experiments in *S. cerevisiae* have shown that Scm3 facilitates CENP-A^{Cse4} nucleosome reconstitution on AT-rich sequences via a conserved AT-hook DNA binding domain that binds non-specifically to AT-rich sequences (Xiao et al., 2011). Both *S. cerevisiae* and *Neurospora crassa* Scm3 proteins contain a predicted AT-hook domain, but this domain is absent from *S. pombe* Scm3 (Aravind et al., 2007). Although Scm3 is a CENP-A^{Cnp1} chaperone, it is possible that it interacts with gap DNA to localise to centromeres.

Despite the absence of an AT-hook S. pombe Scm3 may still associate with the AT-rich DNA between CENP-A^{Cnp1} nucleosomes and such binding might provide a mechanism for targeting CENP-A^{Cnp1} to the centromere. To investigate this possibility ChIP-seq was performed to map Scm3 association across centromeres. The enrichment pattern of Scm3 and CENP-A^{Cnp1} were found to be very similar within all three centromeres. Scm3 was enriched at specific peaks within the

centromere central domains (Figure 5.12). However, Scm3 was also enriched in the gaps between these peaks, although at a relatively low level. In total 42 Scm3 peaks were detected in all three centromeres compare to the 54 CENP-A^{Cnp1} peaks previously identified (Lando et al., 2012). Comparative analyses revealed a significant overlap in the position of these peaks between the two data sets (Figure 5.13). 30 of the peaks were found to be shared (within a 30 bp cut-off limit) between the two data sets. Thus Scm3 localises to the same positions as CENP-A^{Cnp1} nucleosomes and it does not appear to be preferentially associated with the AT-rich gaps that are detected between CENP-A^{Cnp1} nucleosomes.

Scm3 IP/IN ratio coverage was strongly correlated with the published CENP-A^{Cnp1} data distribution (Lando et al., 2012) at 10,000 random sites within the central domains of centromeres (τ=0.77; Figure 5.14A). However, the data deviates from linearity at sites of low CENP-A^{Cnp1} occupancy where moderate levels of Scm3 enrichment were detectable. Thus, at the 10,000 random positions genome-wide that were tested only a weak correlation between Scm3 and CENP-A^{Cnp1} (τ=0.32; Figure 5.14B) was evident. This was mostly driven by sites with the lowest and highest levels of coverage. The input (IN) data were analysed similarly for comparison to see whether extraction and sequencing biases may have affected the correlations observed. CENP-A^{Cnp1} and Scm3 IN coverage at 10,000 randomly chosen positions within the three central domain regions showed a weak correlation (τ=0.24; Figure 5.14C). At 10,000 random sites selected across the genome there was only a weak correlation between CENP-A^{Cnp1} and Scm3 IN coverage (τ=0.22; Figure 5.14D). It is likely that this slight bias shared between the two data sets has a minimal effect given the high levels of enrichment and unbiased reads resulting from ChIP-seq of Scm3.

5.2.6. Condensin localises at the central domain with no clear peaks

All tested centromeric and kinetochore proteins localised to the same sites as CENP-A^{Cnp1} within the central domain or lacked clear peaks. No proteins could be mapped to the regions between CENP-A^{Cnp1} nucleosomes, but there are other proteins enriched at the centromere that could be tested. One such candidate is condensin. Condensin is involved in the structural organisation of condensed chromatin and ChIP-seq of condensin found it is highly enriched at fission yeast centromeres (Kim et al., 2014). The published condensin ChIP-seq data was re-

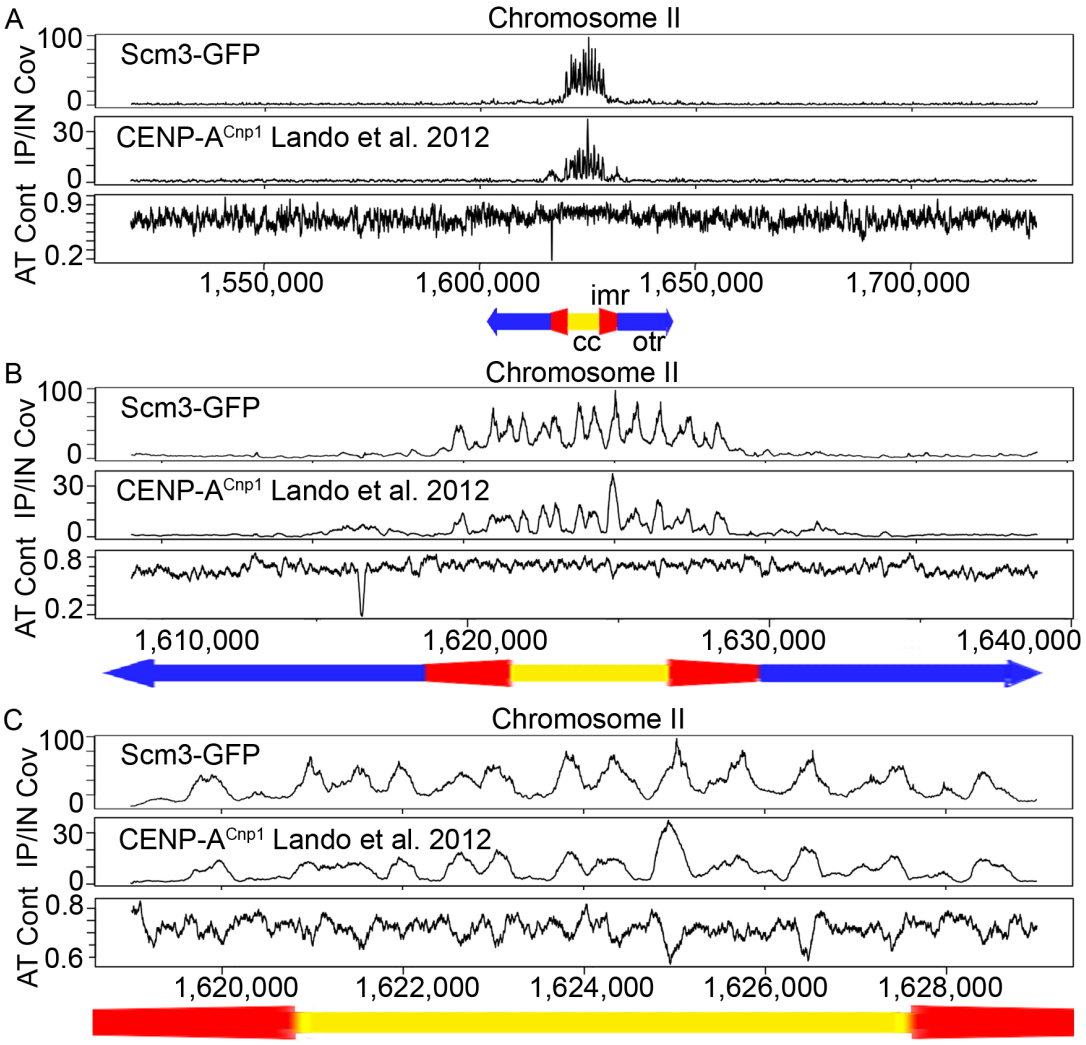


Figure 5.12. CENP-A^{Cnp1} and Scm3 are enriched at the same peaks within centromeres. CENP-A^{Cnp1} (Lando et al. 2012) and Scm3-GFP IP/IN coverage are plotted across the the chromosome II centromere, including (A) 100 kb, (B) 10 kb, and (C) 0 kb around the central domain, as a representative example of centromeric enrichment.

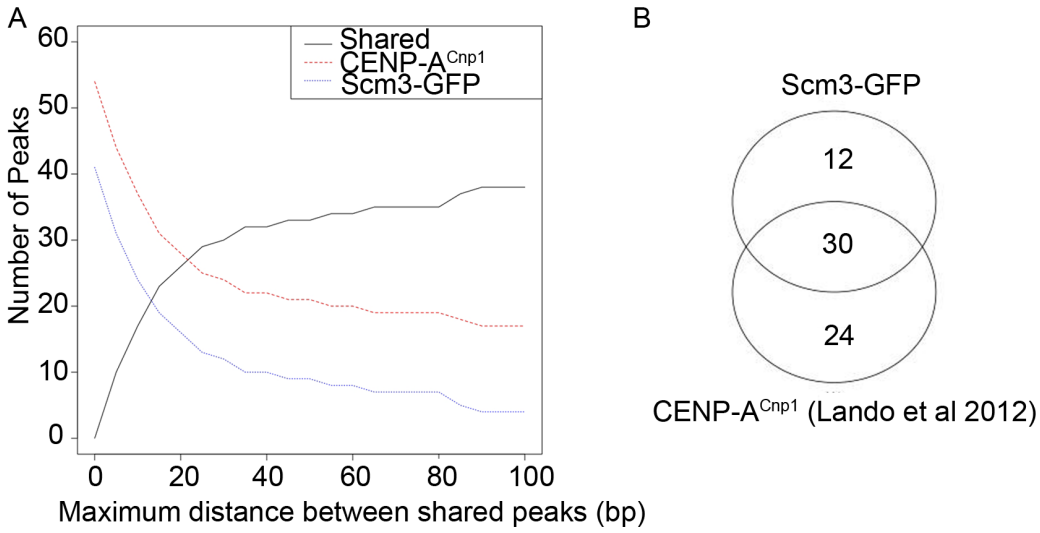


Figure 5.13. CENP-A^{Cnp1} and Scm3-GFP are enriched at the same peaks within centromeres. (A) Scm3-GFP and published CENP-A^{Cnp1} data (Lando et al. 2012) are enriched for peaks within centromeres that are shared between data sets. (B) If shared peaks are defined as being within 30 bp then 30 of the peaks overlap between the two data sets.

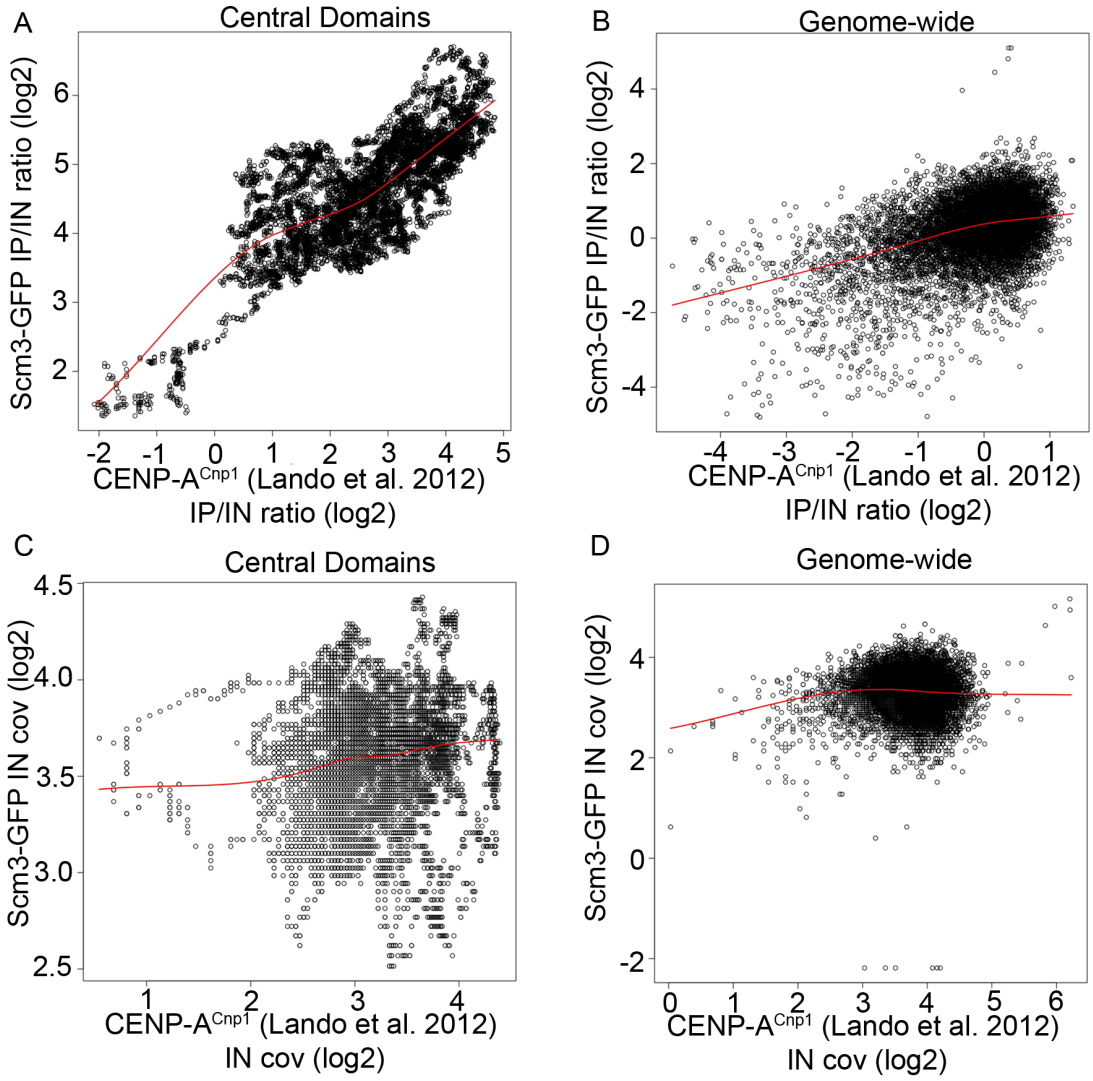


Figure 5.14. Scm3 is not preferentially enriched at AT-rich sequences between CENP-A^{Cnp1} nucleosomes. (A) There is a strong correlation between the IP/IN ratio of Scm3-GFP and the previously published CENP-A^{Cnp1} ChIP-seq data (Lando et al. 2012) at 10,000 random positions within the central domains (τ=0.77). (B) At 10,000 random sites genome-wide there is only a weak correlation between Scm3-GFP and CENP-A coverage, which is driven primarily by sites with low coverage (τ=0.32). Sequencing bias is unlikely to explain the relationship between Scm3-GFP and CENP-A^{Cnp1}. (C) There is a weak correlation in the IN coverage of Scm3-GFP and the previously published CENP-A ChIP-seq data (Lando et al. 2012) at 10,000 random positions within the central domains (τ=0.24). (D) At 10,000 random sites genome-wide there is only a weak correlation between Scm3-GFP and CENP-A^{Cnp1} IN coverage, which is driven primarily by sites with extremely low or high coverage (τ =0.22).

analysed and compared to the CENP-A^{Cnp1} ChIP-seq data to determine whether condensin may associate with the regions between CENP-A^{Cnp1} nucleosomes.

ChIP-seq of the FLAG-tagged condensin component Cut3 has been previously published. These data were generated following a cdc25-22 block/release to synchronise cells in late G2 and analyse Cut3-FLAG distribution through the cell cycle in approximately G2, metaphase, anaphase, and G1/S (Kim et al., 2014). The cell cycle phases given are estimates based on the time after release from the cell cycle block and are not precisely defined. This data was re-analysed in the same way as the CENP-A^{Cnp1} and kinetochore ChIP-seq data using our pipeline. As previously shown (Kim et al., 2014) Cut3-FLAG was enriched over the central domain during metaphase and anaphase, but not during G2 or G1/S (Figure 5.15). The resolution of the Cut3-FLAG condensin data is lower than our centromere and kinetochore protein ChIP-seq data, making peaking calling and thus comparative analyses of enrichment profiles difficult. The problem in peak calling with the Cut3-FLAG data set arises from the fact that a larger average chromatin fragment size (~450 bp) was used to generate these ChIP-seq data and only a single end sequencing was done (Kim et al., 2014). Consequently, the precise size and location of any individual fragment could not be determined and post-sequencing size selection could not be performed to improve resolution. Thus, the localisation of condensin relative to CENP-A^{Cnp1} nucleosomes remains an open question and would require additional ChIP-seq analyses using our methodology.

5.2.7. Microtubules and kinetochore components appear to associate with all primary CENP-A^{Cnp1} nucleosomes

CENP-A^{Cnp1} nucleosomes are present at fission yeast centromeres in excess relative to the number of microtubules that contact each kinetochore during mitosis (Ding et al., 1993; Lando et al., 2012). It is therefore evident that a simple repeat unit based on a one CENP-A^{Cnp1} nucleosome per MT attachment sites, as seen in S. cerevisiae, is not present at fission yeast centromeres. However, this observation opens up numerous possibilities for how CENP-A^{Cnp1} nucleosomes may interact with the kinetochore and microtubules. One, a few, or many CENP-A^{Cnp1} nucleosomes may contact kinetochore components. Such interactions may involve all CENP-A^{Cnp1} nucleosomes within a single centromere or there may be a particular subset of

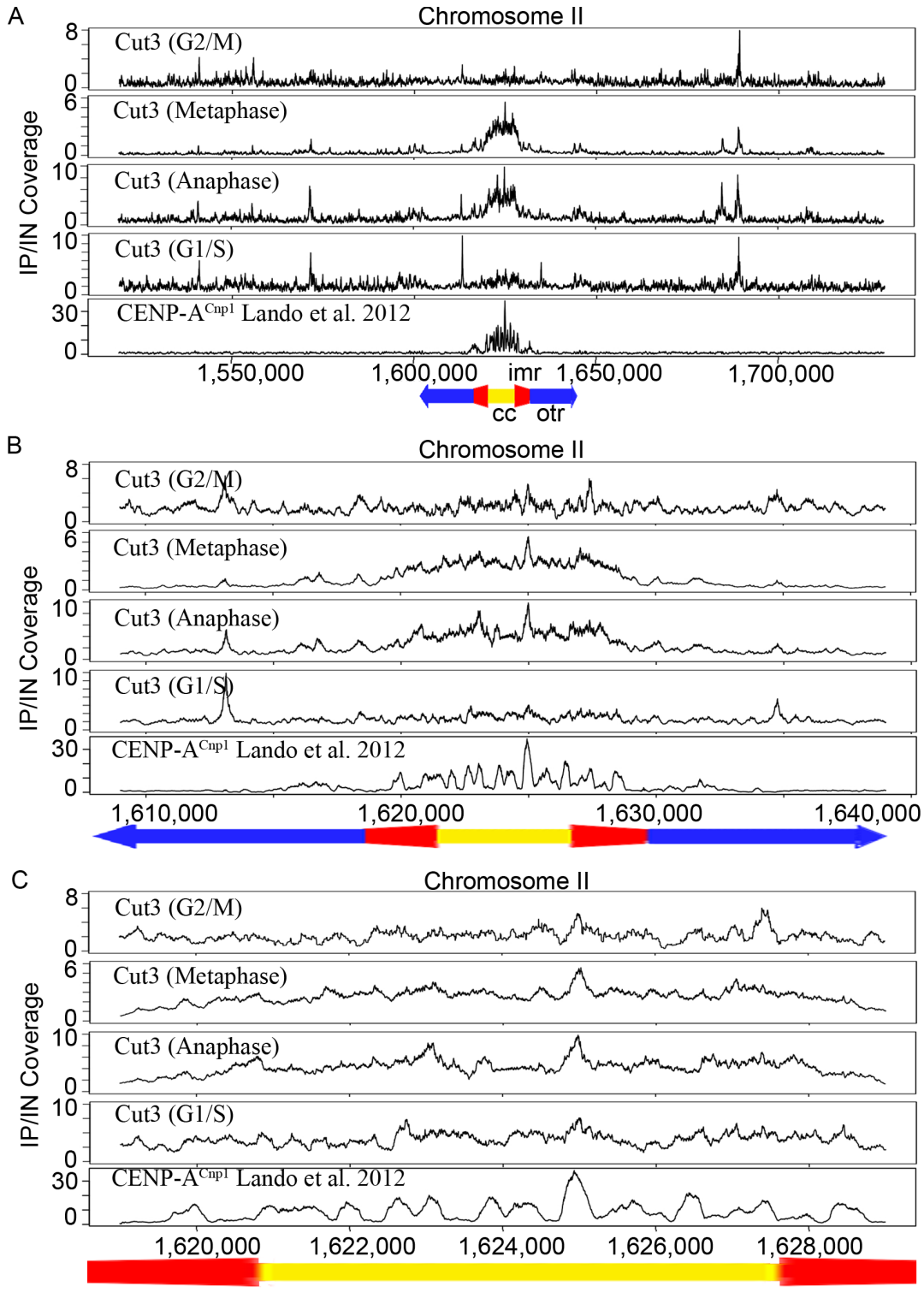


Figure 5.15. Condensin is enriched at centromeres during mitosis. Enrichment of condensin component Cut3 (data from Kim et al., 2014) is shown at the the chromosome II centromere, including (A) 100 kb, (B) 10 kb, and (C) 0 kb of flanking sequences. During metaphase and anaphase Cut3 is enriched at the central domain, but it is not enriched at discrete peaks.

elite CENP-A^{Cnp1} nucleosomes that act as a platform for kinetochore and microtubule attachment site assembly (Figure 5.1B). The Ndc80 complex connects kinetochore components with microtubule. Ndc80-GFP ChIP-seq data was analysed in depth to determine if particular CENP-A^{Cnp1} nucleosomes exhibit more frequent contacts with this outer kinetochore component and thus kinetochore microtubules.

Ndc80 is an outer-kinetochore component in the Ndc80-complex that directly interacts with spindle microtubules (Takeuchi and Fukagawa, 2012). Thus mapping its interactions relative to the underlying sequence will reflect protein-protein crosslinks with kinetochore components that reside in closer proximity to centromeric DNA. Despite being an outer-kinetochore component that is not directly associated with DNA, ChIP-seq reveals that Ndc80-GFP is highly enriched over centromeres (Figure 5.16). Furthermore, the Ndc80-GFP distribution profile exhibits discrete peaks within the central domain of the chromosome II centromere with low coverage obtained between these peaks (Figure 5.16A), but differences observed at other centromeres are discussed below. In total 26 Ndc80-GFP peaks within all centromere central domains were found to coincide (within a 30 bp limit) between the Ndc80-GFP and CENP-A^{Cnp1} data sets, with a further 29 and 28 peaks being specific to each data set, respectively (Figure 5.17). The relationship between the Ndc80-GFP and CENP-A^{Cnp1} enrichment profiles was found to differ between centromeres. While the chromosome II centromere shows discrete peaks with most peaks shared, centromeres on chromosomes I and III showed less well-defined Ndc80-GFP enrichment and the pattern did not correlate as well with the CENP-A^{Cnp1} profile (Figure 5.18). There are a few Ndc80-GFP peaks that do not appear to correlate with CENP-A^{Cnp1} peaks, but the majority of peaks do occur at the same sites so the significance of these additional peaks is unclear. This lack of correlation may be due to the fact that centromeres on chromosomes I and III share a large region of homology so that enrichment profiles obtained are actually an average of two sequences over these shared regions. However, strong underlying sequence characteristics influence CENP-A^{Cnp1} occupancy and differences between centromeres were not observed for any other ChIP-seq data sets. Therefore, a difference in CENP-A^{Cnp1} occupancy between the two regions of shared homology at the centromeres of chromosomes I and III is most likely not responsible for their less distinct Ndc80-GFP profile. At this time the cause of this difference in the

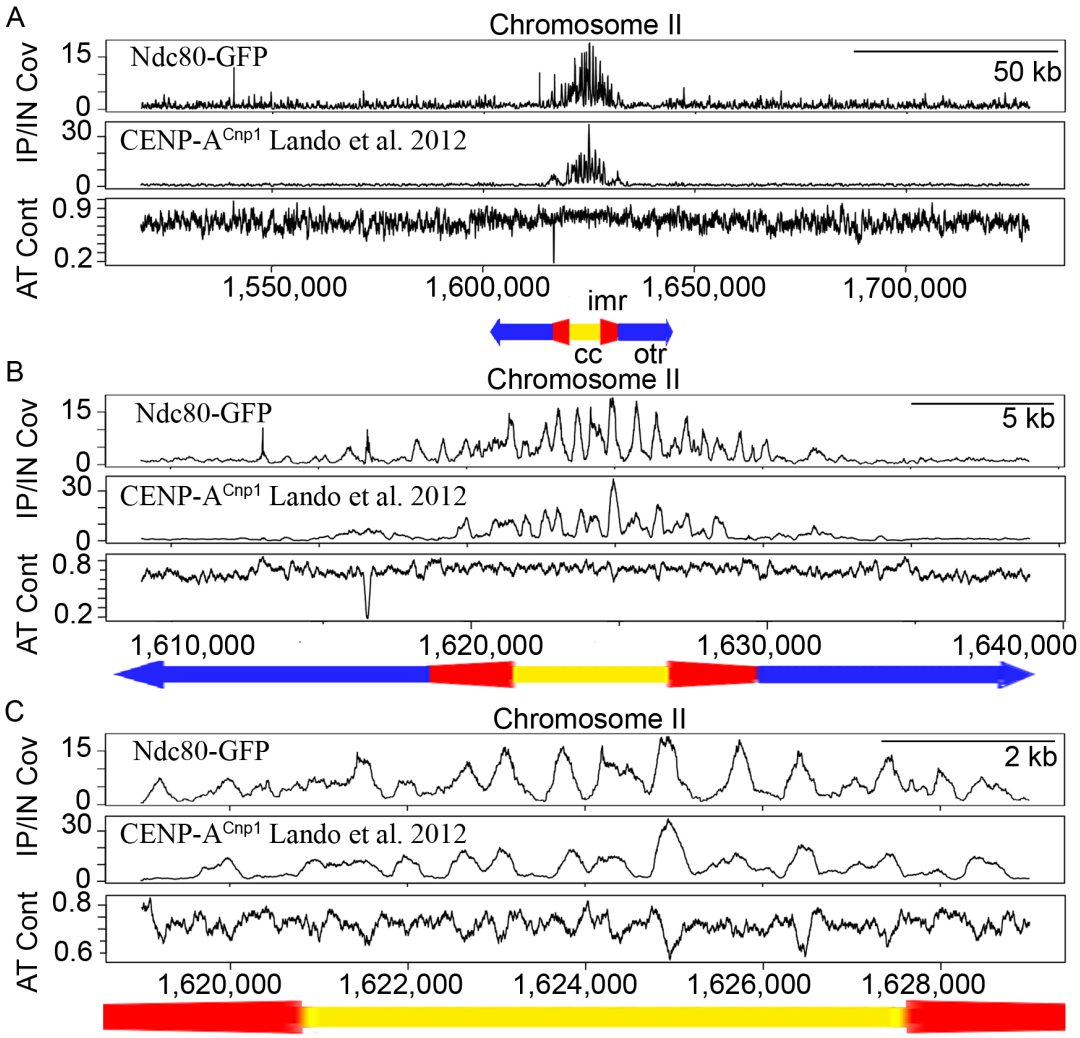


Figure 5.16. CENP-A^{Cnp1} and Ndc80 enrichment within centromeres is correlated. CENP-A^{Cnp1} (Lando et al. 2012) and Ndc80-GFP IP/IN coverage are plotted across the chromosome II centromere, including (A) 100 kb, (B) 10 kb, and (C) 0 kb around the central domain.

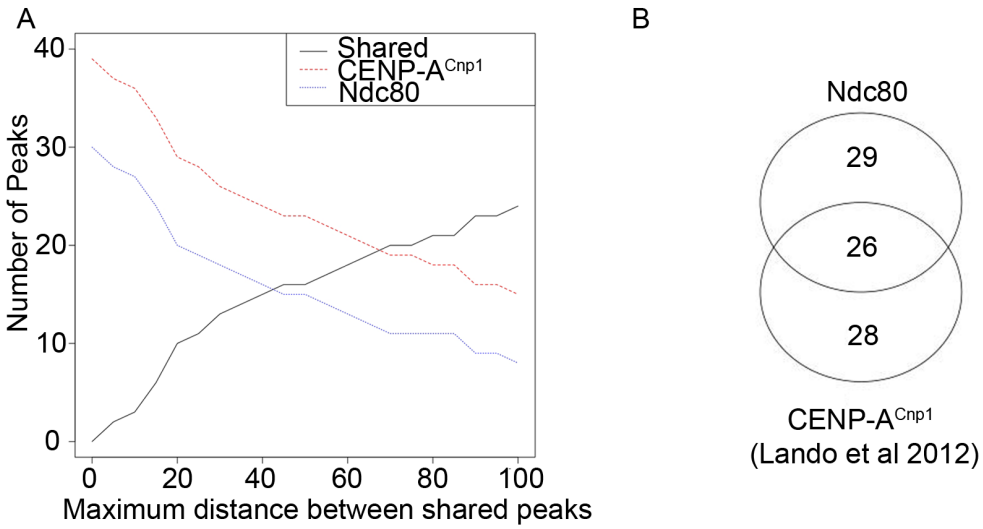


Figure 5.17. All CENP-A^{Cnp1} nucleosomes potentially interact with the kinetochore. (A) CENP-A^{Cnp1} (Lando et al. 2012) and Ndc80-GFP include a large number of shared enrichment peaks within centromeres. (B) If shared peaks are defined as being within 30 bp then 26 of the peaks overlap between the two data sets.

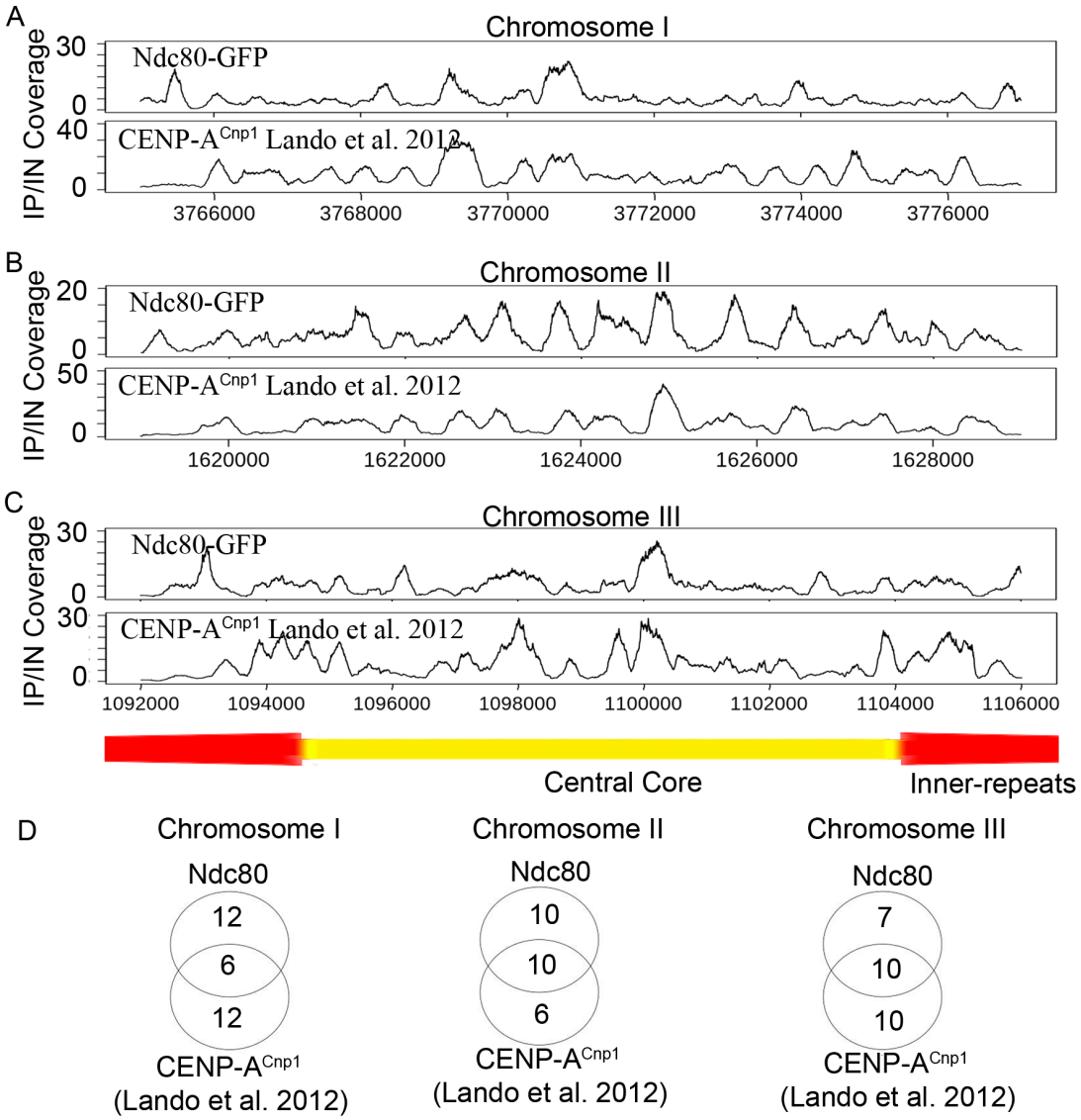


Figure 5.18. Ndc80 shows different enrichment patterns relative to CENP-A^{Cnp1} within different centromeres. Enrichment of Ndc80-GFP and published CENP-A^{Cnp1} data (Lando et al. 2012) are shown at the centromere central domains of chromosomes (A) I, (B) II, and (C) III. (D) Venn diagrams indicate the number of peaks that are specific to each sample and the number of peaks shared between samples (shared peaks within 30 bp). Ndc80-GFP shows strongly enriched discrete peaks on centromere of chromosome II that overlap with CENP-A^{Cnp1} peaks, but within centromeres of chromosomes I and III Ndc80-GFP enrichment overlap less with CENP-A^{Cnp1} peaks.

Ndc80-GFP pattern between different centromeres remains unknown, but may be related to the relatively lower enrichment of Ndc80-GFP at centromeres.
The IP/IN ratio of Ndc80-GFP is moderately correlated with CENP-A^{Cnp1} at 10,000 random sites within the central domains of all three centromeres (τ=0.56; Figure 5.19A). At 10,000 random sites genome-wide there is a weak correlation between Ndc80-GFP and CENP-A^{Cnp1} enrichment (τ=0.26; Figure 5.19B). The IN coverage was also analysed to determine if bias could affect the observed correlations. There was a moderate correlation between Ndc80 and CENP-A^{Cnp1} IN coverage at 10,000 random sites selected within the centromere central domains (τ=0.50; Figure 5.19C). At 10,000 random sites selected genome-wide there was only a weak correlation between Ndc80-GFP and CENP-A^{Cnp1} enrichment (τ=0.26; Figure 5.19D). The relatively similar correlations between CENP-A^{Cnp1} and Ndc80-GFP observed for the IP/IN and the IN coverage are surprising, but the patterns are different. The correlation between Ndc80-GFP and CENP-A^{Cnp1} IN coverage within the central domains raises the concern that bias might be affecting the correlation. However, Ndc80 IN coverage is relatively uniform within the central domains so the small bias observed must result from differences due to the relatively low coverage obtained. In contrast, the CENP-A^{Cnp1} data is known to have a more significant bias and this can be seen in the wider range of IN coverage values within the central domains. The Ndc80-GFP data does not show an IN coverage bias relative to underlying AT-content so it is likely that the data accurately reflects Ndc80-GFP enrichment. Additionally, the correlation observed in the IN is driven primarily by low coverage regions.

Despite the differences in that pattern of Ndc80-GFP peaks between centromeres, multiple Ndc80-GFP peaks coincide with CENP-A^{Cnp1} peaks within all three centromeres and ChIP-seq of the kinetochore proteins more proximal to DNA (CENP-C^{Cnp3} and CENP-T^{Cnp20}) suggest that most CENP-A^{Cnp1} nucleosomes contact or are in close proximity to the kinetochore. Thus, kinetochore assembly and microtubule attachment is not strictly limited to a subset of elite CENP-A^{Cnp1} nucleosomes per centromere. All CENP-A^{Cnp1} nucleosome occupied sites appear to associate with the kinetochore components relatively equivalently. However, ChIP experiments provide an average profile of the many cells present in the population. It is possible that CENP-A^{Cnp1} nucleosomes and associated kinetochore occupy a different selection of sites in different cells. Kinetochore components may form

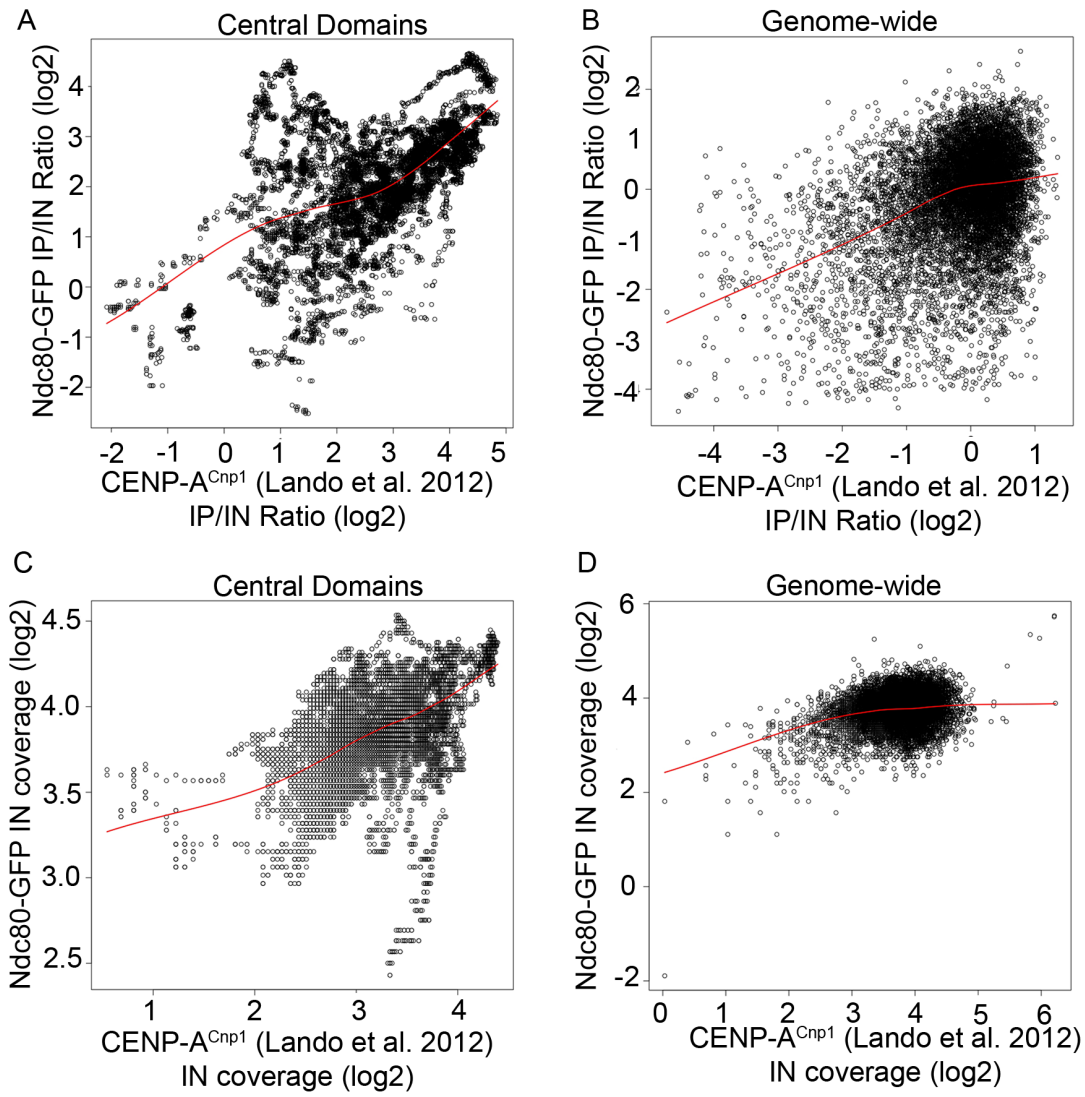


Figure 5.19. Ndc80 is enriched at the same enriched peaks as CENP-A^{Cnp1} nucleosomes. (A) There is a moderate correlation between the IP/IN ratio of Ndc80-GFP and the previously published CENP-A^{Cnp1} ChIP-Seq data (Lando et al. 2012) at 10,000 random positions within the central domains (τ=0.56). (B) At 10,000 random sites genome-wide there is only a weak correlation between Ndc80-GFP and CENP-A^{Cnp1} coverage, which is driven primarily by sites with low coverage (τ=0.26). Influence of sequencing and extraction bias on data sets. (C) There is a moderate correlation in the IN coverage of Ndc80-GFP and the previously published CENP-A^{Cnp1} ChIP-Seq data (Lando et al. 2012) at 10,000 random positions within the central domains (τ=0.50). (D) At 10,000 random sites genome-wide there is only a weak correlation between Ndc80-GFP and CENP-A^{Cnp1} IN coverage, which is driven primarily by sites with extremely low or high coverage (τ=0.26).

individual connections that mediate interactions between a single microtubule and a single underlying nucleosome. Alternatively, the kinetochore may form a single large organised structure that can interact with multiple microtubules on the outer surface and contact multiple nucleosomes on its inner surface (Dong et al., 2007; Zinkowski et al., 1991). ChIP-seq analyses alone cannot distinguish between these possibilities.

5.2.8. Eic1 and Eic2 association profiles are consistent with their role as CENP-A^{Cnp1} recruitment or stabilising factors

Fission yeast lacks Mis18BP1 that is responsible for the recruitment of the CENP-A loading factors Mis18α/β to vertebrate centromeres. Eic1 and Eic2 were identified as two additional proteins involved in recruiting the Mis16/Mis18 complex to fission yeast centromeres and consequently for CENP-A^{Cnp1} incorporation (Hayashi et al., 2014; Subramanian et al., 2014). Eic1 was found to be essential and required for Mis16, Mis18, and Scm3 localisation to centromeres. Eic2 was not essential, but it appeared to play an independent role in contributing to Mis18 function. The distributions of Eic1-GFP and Eic2-GFP were also examined by ChIP-seq as part of one of these studies to determine if their localisation was consistent with their function as potential centromere specific recruitment factors for the Mis16-Mis18 complex (Subramanian et al., 2014).

Eic1-GFP and Eic2-GFP were highly enriched over the central domain regions, but were not enriched on the outer-repeats or at other locations across the genome (Figure 5.20). Enrichment was compared to the Mis16, Mis18, and Scm3 ChIP-seq profiles within the centromere central domains to determine if their enrichment supports a role as Mis16-Mis18 complex recruitment factors. At 10,000 random sites from the three central domains Eic1-GFP IP/IN ratios were strongly correlated with Scm3-GFP (τ=0.88) and Mis18-GFP (τ=0.82) IP/IN ratios but weakly correlate with Mis16-GFP IP/IN ratios (τ=0.37; Figure 5.21), which is not surprising given that Mis16-GFP is not enriched at discrete peaks within the central domains. In contrast, Eic1-GFP IP/IN ratios were weakly correlated with Scm3-GFP (τ=0.23), Mis16-GFP (τ=0.069), and Mis18-GFP (τ=0.24) IP/IN ratios at 10,000 random sites genome-wide. Similarly, within the three central domains Eic2-GFP IP/IN ratios were strongly correlated with Scm3-GFP (τ=0.84) and Mis18-GFP (τ=0.85) IP/IN ratios but weakly correlate with Mis16-GFP IP/IN ratios (τ=0.37; Figure 5.22). In contrast, Eic2-GFP

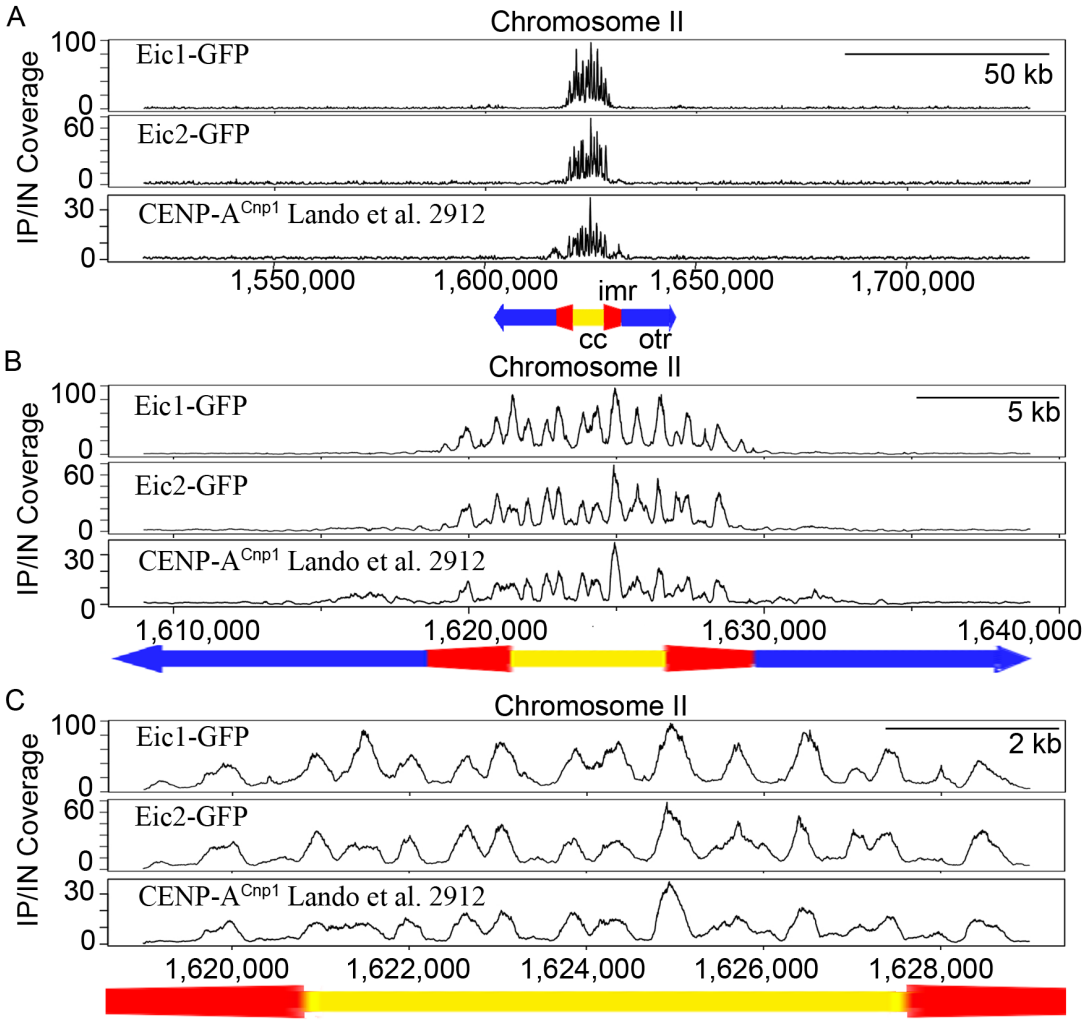


Figure 5.20. Eic1 and Eic2 are enriched at peaks within the central domains. ChIP-seq of Eic1-GFP and Eic2-GFP, as well as the published CENP-A^{Cnp1} data (Lando et al. 2012), is shown at the chromosome II centromere, including (A) 100 kb, (B) 10 kb, and (C) 0 kb of flanking sequence around the central domain, as a representative example of centromeric enrichment. Both Eic1-GFP and Eic2-GFP are highly enriched specifically at the central domains with no enrichment at the centromeric outer-repeats or other genomic loci. Eic1-GFP and Eic2-GFP map to well-defined peaks within the central domains.

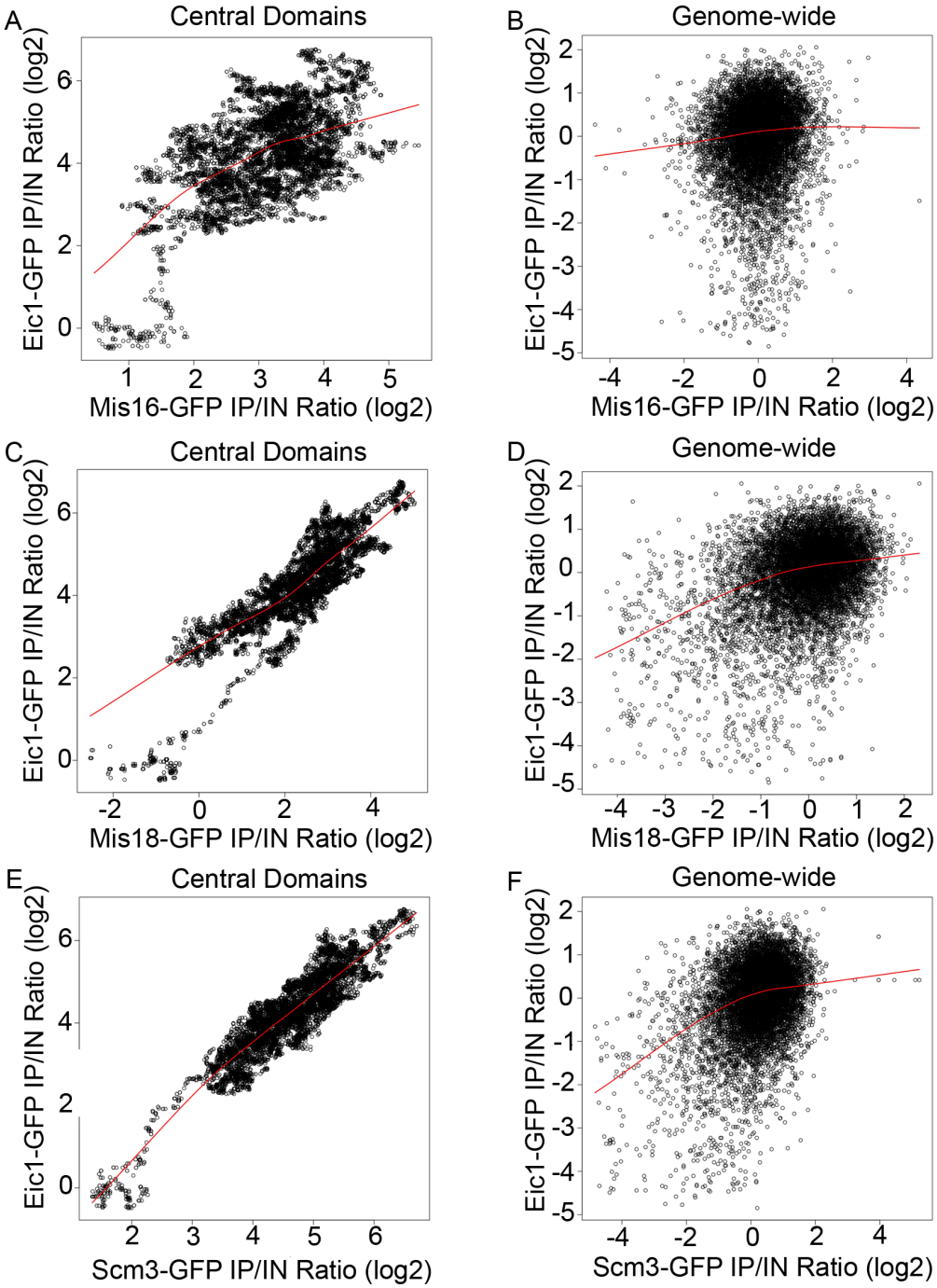


Figure 5.21. Eic1 is a candidate Mis16-Mis18 recruitment or stabilising factor. Correlations tested between IP/IN ratios at 10,000 random sites within the central domains and genome-wide. Eic1-GFP correlates weakly with Mis16-GFP within the central domains (A; τ=0.37) and genome-wide (B; τ=0.069). Eic1-GFP correlates strongly with Mis18-GFP within the central domains (C; τ=0.82) but correlates weakly genome-wide (D; τ=0.24). Eic1-GFP correlates strongly with Scm3-GFP within the central domains (E; τ=0.88) but correlates weakly genome-wide (F; τ =0.23).

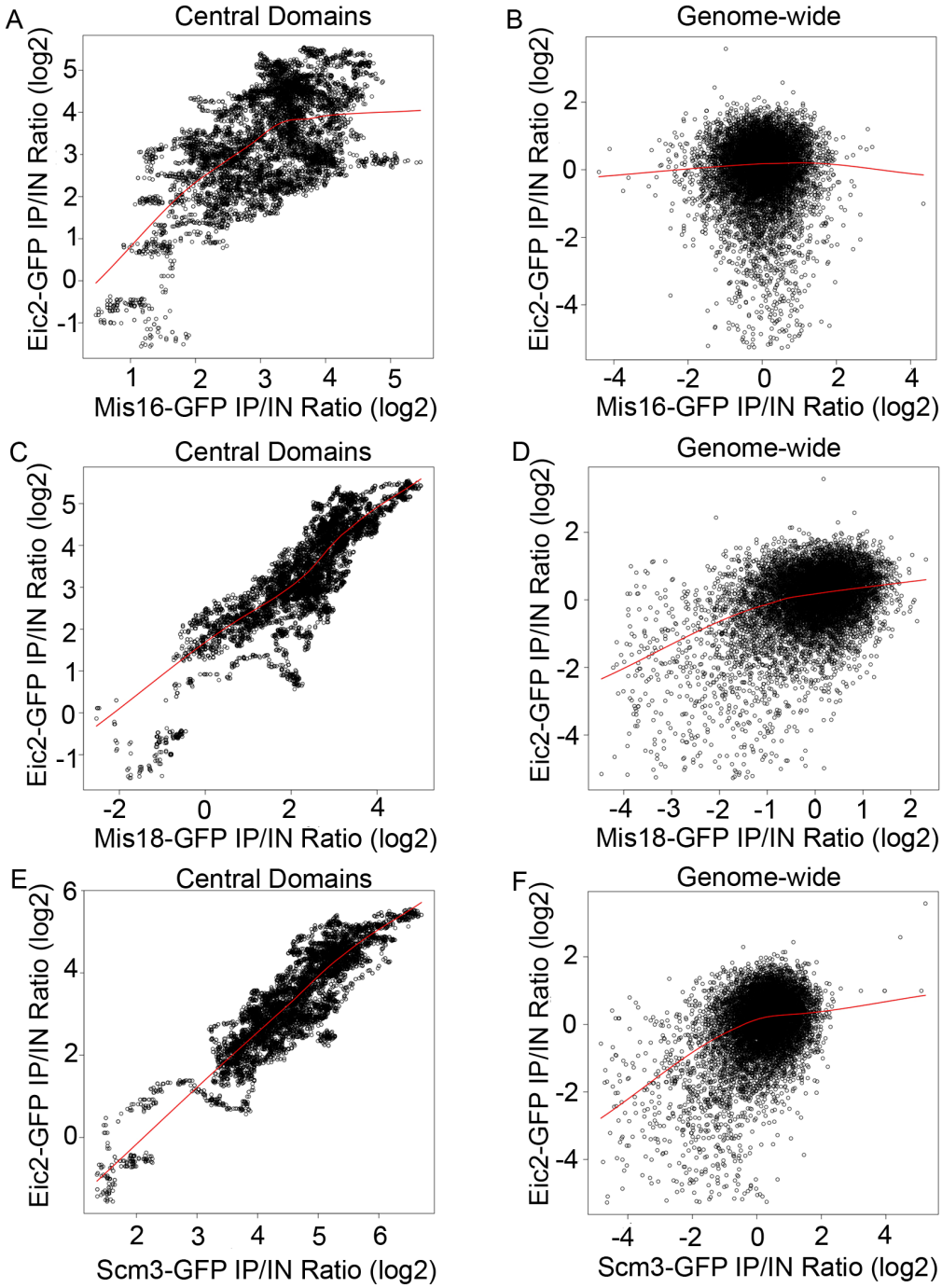


Figure 5.22. Eic2 is a candidate Mis16-Mis18 recruitment or stabilisation factor. Correlations tested between IP/IN ratios at 10,000 random sites within the central domains and genome-wide. Eic2-GFP correlates weakly with Mis16-GFP within the central domains (A; τ=0.37) and genome-wide (B; τ=0.018). Eic2-GFP correlates strongly with Mis18-GFP within the central domains (C; τ=0.85) but correlates weakly genome-wide (D; τ=0.24). Eic2-GFP correlates strongly with Scm3-GFP within the central domains (E; τ=0.84) but correlates weakly genome-wide (F; τ =0.27).

IP/IN ratios were weakly correlated with Scm3-GFP (τ=0.27), Mis16-GFP (τ=0.018), and Mis18-GFP (τ=0.24) IP/IN ratios genome-wide. This weak correlation genome-wide was primarily due to regions of low coverage and likely represents influences of chromatin structure and extraction efficiency on sequencing since these samples were not affected by any sequencing bias associated with AT-content. Prominent well-defined peaks were evident within the profiles of both Eic1-GFP and Eic2-GFP enrichment across the central domains. This localisation supports the role of Eic1 and Eic2 as Mis16-Mis18 complex recruitment or stabilising factors.

5.3 Discussion

The new CENP-A^{Cnp1} ChIP-seq data generated in this study (Chapter 3) suggested that the CENP-A^{Cnp1} occupancy pattern may not be as well-defined as previously reported (Lando et al., 2012). There may be some secondary sites that are occupied by CENP-A^{Cnp1} nucleosome and positioning may not be as tightly constrained. However, ChIP-seq of multiple centromere-kinetochore proteins suggests that interactions between CENP-A^{Cnp1} and the kinetochore take place at the most highly positioned sites that are defined by associated DNA sequence features. Transient or unstable CENP-A^{Cnp1} nucleosomes that occupy other sites are therefore either unlikely to interact with kinetochore components or to be present at a low frequency in the cell population. The profiles obtained for multiple proteins support the hypothesis that the more dispersed CENP-A^{Cnp1} profile generated in the new dataset resulted from experimental limitations, for example cross-linking between CENP-A^{Cnp1} and other complexes or DNA, and that CENP-A^{Cnp1} is in fact highly positioned by asymmetric AT-rich sequences.

The data generated allow a number of conclusions can be drawn with respect to the overall organisation of centromere-associated proteins. CENP-A^{Cnp1} primarily occupies discrete sites within the central domains that act as the sites of interaction with kinetochore components. Ndc80, CENP-T^{Cnp20}, and CENP-C^{Cnp3} enrichment peaks overlap with all CENP-A^{Cnp1} occupied sites, indicating that all CENP-A^{Cnp1} nucleosomes interact with or are within close proximity to kinetochore components. Since ChIP-seq is performed on a population of cells it is possible that not all sites are occupied in any given cell in a population. Thus different CENP-A^{Cnp1} nucleosomes may direct interactions with the kinetochore in different cells. However, all the positioned CENP-A^{Cnp1} occupied sites appear to interact with kinetochore

proteins, including the Ndc80 complex, and by inference with microtubules (Figure 5.1B). The data generated rule out the possibility that a few elite CENP-A^{Cnp1} nucleosomes provide the predominant platforms for connecting centromeric DNA with the kinetochore components and microtubules.

Given its distance from the underlying centromeric DNA it is somewhat surprising that the outer kinetochore component Ndc80 can be mapped with such high resolution since this necessitates multiple protein-protein crosslinks between several kinetochore proteins. Crosslinking must result in extensive crosslinking between a network of CENP-A^{Cnp1} and kinetochore components so that any DNA sequence underlying CENP-A^{Cnp1} and the kinetochore is enriched regardless of its actual proximity to that particular kinetochore component. This means that ChIP-seq of kinetochore proteins provides insight into all interactions between CENP-A^{Cnp1} and all kinetochore proteins regardless of whether its association occurs via CENP-T^{Cnp20} or CENP-C^{Cnp3} and the Mis12 complex. Therefore, ChIP-seq of CENP-T^{Cnp20} or CENP-C^{Cnp3} cannot distinguish which protein is associated with a given enriched sequence since the crosslinking network will result in enrichment of all sequences in both CENP-T^{Cnp20} and CENP-C^{Cnp3} ChIPs. However, the crosslinking network does not affect the observation that all CENP-A^{Cnp1} nucleosomes interact with the kinetochore since the CENP-A^{Cnp1} profile does not rely on differentiating the path via which interactions with kinetochore proteins takes place.

Despite the high quality and resolution of the ChIP-seq data generated, no candidate protein was found to be preferentially enriched on the sequences residing between CENP-A^{Cnp1} occupied sites. Additionally, CENP-A^{Cnp1} may not be present at the centromeres in the required numbers to completely occupy all sites, although this conclusion is based on the use of a tagged CENP-A^{Cnp1} with compromised functionality (Lando et al., 2012). However, it remains possible that another protein or protein complex occupies the same sites as CENP-A^{Cnp1} in different cells so that when CENP-A is not present that site is occupied by this alternative protein(s). Whether a distinct protein or complex also occupies the gaps that are detected between CENP-A^{Cnp1} nucleosome remains unclear. One possibility was the CENP-T^{Cnp20}/W/S/X complex, which has been proposed to form a nucleosome-like structure in vertebrates (Nishino et al., 2012). However, CENP-T^{Cnp20} did not localise to the regions between CENP-A^{Cnp1} nucleosomes, but instead appeared to occupy

the same sites as CENP-A^{Cnp1}. A more extreme possibility is that the regions not occupied by CENP-A^{Cnp1} could exist as free DNA. While these analyses have taken a candidate approach to investigating what proteins may be associated with the gaps between CENP-A^{Cnp1} nucleosomes, it would also be possible to try to identify such proteins through other methods. One possibility is to digest chromatin to mono- and di-nucleosomes and analyse what proteins are differentially associated with di-nucleosomes by mass spectrometry.

Several candidates remain to be tested for their co-occupancy of CENP-A^{Cnp1} sites or as proteins that preferentially reside between CENP-A^{Cnp1} peaks. However, the two most promising candidates have also thus far proven to be the most difficult to obtain high quality ChIP-seq data for. The most obvious candidate for a complex that could occupy the same sites as CENP-A^{Cnp1} nucleosomes is an H3 nucleosome. There is a strong argument for this possibility: H3 nucleosomes may be incorporated into centromeric chromatin following DNA synthesis during S phase to act as a placeholder until CENP-A deposition later in the cell cycle (Allshire and Karpen, 2008). Indeed, in human cells the histone H3 variant H3.3 is incorporated into centromeric chromatin during S phase and replaced by CENP-A during G1 (Dunleavy et al., 2011). The deposition of new CENP-A^{Cnp1} occurs primarily during G2 in *S. pombe* so it is possible that H3 nucleosomes act as a placeholder following centromere replication in early S phase through to CENP-A deposition in mid-late G2 (Lando et al., 2012). Since there is a precedent for H3 nucleosomes being temporary placeholders for CENP-A, a natural extension of this observation is that additional H3 nucleosomes may also occupy sites within the central domain that remain unoccupied by CENP-A^{Cnp1}. This possibility could not be satisfactorily investigated in this study because the enrichment of centromeric central domain sequences following an H3 ChIP was found to be too low. It remains unclear whether this low yield is because H3 nucleosomes are actually extremely rare within central domain chromatin or whether it is due to other experimental limitations, such as extraction efficiency. Until this limitation can be addressed it will not be possible to evaluate whether H3 nucleosomes occupy sites within the central domain throughout the cell cycle.

The other promising candidate for a protein complex that occupies sites within the central domain of fission yeast centromeres is the origin recognition complex (ORC).

DNA replication is initiated during S-phase at origins of replication that are bound by ORC, a six subunit protein complex consisting of Orc1-6 (Cayrou et al., 2012; Li and Stillman, 2012). The ORC proteins were originally identified in *S. cerevisiae* and orthologs have been widely described among characterized eukaryotes (Dhar and Dutta, 2000; Foss et al., 1993; Grallert and Nurse, 1996; Kong et al., 2003; McNairn et al., 2005; Zisimopoulou et al., 1998). Surprisingly, ORC components also localizes within the central domains of fission yeast centromeres where no replication origin activity has been observed (Hayashi et al., 2007; Segurado et al., 2003). ChIP-chip experiments also indicated that the MCM helicases were not present and that the region did not show evidence of origin firing (Hayashi et al., 2007). Direct tests for DNA replication using 2D agarose gels that detect replication intermediates at origins did not detect evidence of active origins within centromeric central domains (Kim et al., 2003). However, subsequent ChIP-chip experiments have provided evidence for both the presence of the MCM proteins and early S phase nucleotide incorporation (Hayano et al., 2012). Thus, the available data on whether the central domain associated ORC functions to promote origin activity remains unresolved. However, fission yeast ORC has previously been proposed to function in silencing since there is an enrichment of AT-rich islands at telomeres, centromeres, and the mating type locus, which are all associated with H3K9-metylation dependent heterochromatin (Segurado et al., 2003).

ORC is already known to function in a number of processes besides replication. ORC is required for the maintenance of heterochromatin structure in mammals and *Drosophila* (Auth et al., 2006; Badugu et al., 2005; Pak et al., 1997; Prasanth et al., 2010; Shareef et al., 2001, 2003). In budding yeast, which lacks H3K9 methylation and canonical heterochromatin components, Orc1 acts as a platform for the silent information regulator (Sir) proteins that transcriptionally silence telomeres and the silent mating type loci (Hickman et al., 2011; Shareef et al., 2003). Given the known functions of ORC outside of replication, it is possible that *S. pombe* ORC has a different function within the central domain of centromeres. Fission yeast ORC binds to asymmetric AT-rich sequences through the multiple N-terminal Orc4 AT-hooks (Chuang and Kelly, 1999; Kong and DePamphilis, 2001; Lee et al., 2001). The gaps between CENP-A^{Cnp1} nucleosomes are enriched for asymmetric AT-rich sequences (Chapter 3) and thus ORC is a prime candidate for a complex that is predicted to bind to these sequences that reside between CENP-A^{Cnp1} nucleosomes. However,

the published ChIP-chip data of ORC components do not have the spatial resolution to determine precisely where ORC is localised within centromeres. Nevertheless, the enrichment of ORC within the central domain regions indicates that ORC occupies either the same sites as CENP-A^{Cnp1} nucleosomes or the gaps between CENP-A^{Cnp1} nucleosomes. Therefore, ORC ChIP-seq will provide at least a partial explanation for either the gaps between CENP-A^{Cnp1} nucleosomes or the discrepancy between the number of CENP-A^{Cnp1} nucleosomes and the number of potentially occupied sites. Efforts to apply ChIP-seq to ORC components in this study were unsuccessful because the ChIP samples did not have adequate enrichment for high throughput sequencing. New methodologies with greater spatial resolution and lower experimental background, such as ChIP-exo, are currently being applied to address these questions (M. Shukla, Allshire lab, personal communication; Rhee and Pugh, 2011).

These observations are currently somewhat limited in the extent of the conclusions that can be made, but there are other questions that could potentially be addressed using this methodology. For example, the question of whether a cell cycle specific placeholder occupies CENP-A^{Cnp1} sites at some point during the cell cycle remains important. In addition, it remains to be determined if the profile of CENP-A^{Cnp1} occupancy changes at different sites within centromeres during the cell cycle. These and other related questions are currently being pursued.

Chapter 6: Discussion

This thesis explored the relationship between the organisation of sequence features and occupancy of the histone variant CENP-A^{Cnp1} within the central domains of fission yeast centromeres and at neocentromeres. The studies described here indicate that CENP-A^{Cnp1} nucleosomes at *S. pombe* centromeres are organised so that they occupy discrete positions within the central domains. These nucleosomes appear to be positioned by large gaps of asymmetrical AT-rich sequences. These asymmetrical AT-rich sequences could potentially play a dual role in positioning CENP-A^{Cnp1} nucleosomes and in allowing another complex, possibly the origin recognition complex (ORC), to bind to DNA within these gaps between CENP-A^{Cnp1} nucleosomes. Highly positioned CENP-A^{Cnp1} nucleosomes separated by large gaps of asymmetric AT-rich sequences are also observed at *S. octosporus* centromeres, which are conserved in chromosomal position but have no sequence homology relative to *S. pombe* centromeres. This difference in centromere sequence suggests that features such as asymmetric AT-rich motifs may be a general property of centromeres in fission yeast species. In contrast, *S. pombe* neocentromeres that form in subtelomeric regions lack highly positioned CENP-A^{Cnp1} nucleosomes and the gaps between their CENP-A nucleosomes are not enriched for any sequence motif, indicating that normal centromere associated sequence features may be important for optimal centromere function but are not necessary for centromere function. These analyses highlighted a conserved feature that may promote optimal centromere organisation in fission yeast species, but also reaffirms that there is a large degree of plasticity in centromere organisation that therefore allows for neocentromeres to form on non-centromeric sequences.

CENP-A^{Cnp1} over-expression leads to ectopic incorporation of additional CENP-A^{Cnp1} primarily at regions associated with heterochromatin. There may also be increased CENP-A^{Cnp1} incorporation within central domains and evidence suggests that this could lead to CENP-A^{Cnp1} nucleosomes occupying additional sites that are normally not occupied. This overloading of endogenous centromeres with CENP-A^{Cnp1} could lead to eviction of H3 nucleosomes or other complexes that normally occupy these locations. The location of preferential ectopic CENP-A^{Cnp1} incorporation overlaps with the chromosomal regions where neocentromeres are known to form. However, other processes must influence the extent of the CENP-A^{Cnp1} domain at

neocentromeres since over-expression of CENP-A^{Cnp1} results in its incorporation occurs over a larger domain than that observed at neocentromeres.

Identification of highly positioned CENP-A^{Cnp1} nucleosomes with intervening large gaps instigated a search for proteins that may localise to these gaps. So far candidates for proteins that occupy the gaps between CENP-A^{Cnp1} nucleosomes or occupy the same positions as CENP-A^{Cnp1} nucleosomes have not been verified. H3 nucleosomes remain the most promising candidate for a complex that may occupy the same positions as CENP-A^{Cnp1}, but experimental restrictions due to low coverage within centromeres prevented this possibility from being analysed. One candidate for a complex binding to DNA in the gaps between CENP-A^{Cnp1} nucleosomes was the CENP-T/W/S/X complex, which has been proposed to form a nucleosome-like complex in vertebrates (Nishino et al., 2012), but CENP-T^{Cnp20} was found to be enriched at the same sites as CENP-A^{Cnp1} and can therefore be discarded as a candidate. ORC is a promising candidate for a complex occupying the gaps between CENP-A^{Cnp1} nucleosomes since it is enriched at the central domains and the asymmetric AT-rich sequences in the gaps between CENP-A^{Cnp1} represent preferred binding sites for the Orc4 AT-hooks (Chuang and Kelly, 1999; Hayashi et al., 2007; Kim et al., 2003; Kong and DePamphilis, 2001; Lee et al., 2001; Segurado et al., 2003). However, so far this possibility could not be evaluated because ChIP of ORC components did not result in sufficient enriched at ORC bound sits for processing for ChIP-Seq.

6.1. Neocentromere evolution and 'optimal' centromeres

Neocentromeres formed on human chromosomes were initially observed in a clinical setting due to associated chromosomal abnormalities, particularly those that can be identified in cytogenic screens (Marshall et al., 2008). These chromosomal rearrangements often lead to partial trisomy or tetrasomy due to an inverted duplication, but even chromosomal rearrangements that do not lead to partial aneuploidy can cause phenotypic abnormalities (Marshall et al., 2008). However, human neocentromeres have also been observed in the absence of any chromosomal abnormalities or associated disease (Capozzi et al., 2009; Tyler-Smith et al., 1999). Such neocentromeres were only identified by chance in prenatal screenings due to maternal age and were present for at least two generations prior to their detection so the actual prevalence of this type of neocentromere is entirely

unknown (Capozzi et al., 2009; Tyler-Smith et al., 1999). These neocentromeres form at new sites on non-repetitive sequences that are not homologous to the conventional ancestral centromere, which is epigenetically inactivated on the same chromosome (Marshall et al., 2008). Neocentromere formation has also been induced in chicken cells (Shang et al., 2013) and fission yeast (Ishii et al., 2008; Ogiyama et al., 2013), amongst others, by selective deletion of the canonical centromere and subsequent neocentromere formation on non-homologous sequences.

It is of particular interest to understand how neocentromere location is determined, as this must also relate to the processes that control the identity of canonical centromeres. Certain regions on human chromosomes tend to form neocentromeres more frequently (Marshall et al., 2008), but detailed analyses revealed that neocentromeres form on different sequences within these preferred domains (Alonso et al., 2003). Therefore, it has been proposed that neocentromere formation takes place in an epigenetic manner independent of underlying DNA sequence (Alonso et al., 2003). Alternately, it has been proposed that neocentromeres are in fact re-activated latent centromeres (Dutrillaux, 1979; Sart et al., 1997; Voullaire et al., 1993). The centromeres on human chromosomes with a neocentromeres have been observed to reposition back to sites of the original inactivated ancestral centromeres (Capozzi et al., 2009; Ventura et al., 2004). This process may rely on latent epigenetic cues (Ventura et al., 2004) or alternatively regions with a high rate of chromosomal rearrangements may generally tend to form neocentromeres (Ventura et al., 2003).

Neocentromeres can become fixed within the population and subsequently propagate as a canonical centromere, termed an evolutionary new centromere (ENC). The chicken 5, 27, and Z chromosomes have centromeres that are not associated with repeat elements and have thus been proposed to be ENCs, although the evolutionary history of these chicken centromeres is not known (Shang et al., 2010). Similarly, the horse *Equus caballus* chromosome 11 has an ENC that is not associated with repetitive elements and not found at the same position in other mammals (Wade et al., 2009). Centromeric satellite repeats are absent on horse chromosome 11, but on all other chromosomes CENP-A co-localises with centromeric satellite repeats (Piras et al., 2010). This is supported by analysis of the

zebra, horse, and donkey genomes, which identified the presence of a horse ENC and identified 8 ENCs, 5 being specific to the donkey (Carbone et al., 2006). ENCs have been even more thoroughly studied in primates. Rhesus macaques have 9 ENCs, while humans have 5 ENCs, and all macaque ENCs are associated with satellite repeats (Ventura et al., 2007). This may represent a progression towards evolutionarily mature ENCs that initially reposition onto non-repetitive sequences and later become associated with satellite repeat elements (Ventura et al., 2007).

ENCs have been proposed to share a general evolutionary progression (Dawe & Henikoff, 2006; Malik, 2002). Most ENCs are thought to initially form as neocentromeres on non-repetitive sequences. This is true for human neocentromeres initially observed in a clinical setting as well as ENCs in horses and chickens (Marshall et al., 2008; Shang et al., 2013; Wade et al., 2009). Once formed these non-repetitive ENCs are then thought to gradually accumulate satellite arrays over evolutionary timescales (Ventura et al., 2007). This is observed in macaques where ENCs have become associated with repetitive arrays and the ancestral centromeres have lost the associated repetitive arrays (Ventura et al., 2007). This suggests that there is a general evolution of neocentromeres towards an optimal organisation as they evolve from neocentromeres to ENCs and raises the question of what features contribute to this optimal organisation. A number of ideas have been proposed: satellite repeat length reflects occupancy of a single phased nucleosome per repeat and may contribute to nucleosome packaging; alpha-satellite repeats can recruit CENP-B through their CENP-B box, which is not found at human non-alphoid neocentromeres; and satellite repeats could contribute to kinetochore size and therefore strength, which would be selected for by centromere drive (Dawe and Henikoff, 2006; Malik, 2002).

Indeed, centromere strength in mice has been observed to provide a selective force leading to centromere drive (Chmátal et al., 2014). Therefore, centromere drive likely affects optimal centromere organisation in some species with asymmetric meiosis. Additionally, it has recently been shown that CENP-B stabilises kinetochores in mammalian cells by forming a bridge between CENP-A and CENP-C (Fachinetti et al., 2015). Thus, although neocentromeres are functional, additional features, such as the acquisition of CENP-B binding, may enhance their functionality. However, fission yeasts lack asymmetric meiosis and do not contain a CENP-B

homolog and thus other influences must contribute to optimal centromere organisation in these species.

As discussed in this thesis, there is evidence that fission yeast centromeres reflect an optimal organisation that is not found at neocentromeres. Canonical centromeres in fission yeast are comprised of a CENP-A^{Cnp1} occupied central domain of non-repetitive sequence where kinetochore formation occurs and flanking heterochromatic outer-repeat elements that are each several kb in length (Partridge et al., 2000; Takahashi et al., 1992). Within the central domains of these centromeres CENP-A^{Cnp1} nucleosomes are highly positioned at discrete sites by the presence large gaps of asymmetric AT-rich sequences which disfavour nucleosome occupancy. Neocentromere organisation differs from that of endogenous centromeres, being proximal a single region of heterochromatin (i.e. telomere regions) and lacking highly positioned CENP-A^{Cnp1} and asymmetric AT-rich gaps. These results provide insight into what constitutes an optimal centromere, although this may differ between fission yeast and higher eukaryotes. Fission yeast centromeres do not contain CENP-B boxes and have symmetric meiosis so centromere drive is not a selective force, though heterochromatic pericentric repeats contribute to chromosome cohesion and may thus still be important for chromosome stability (Clarke and Baum, 1990; Nonaka et al., 2002; Watanabe and Nurse, 1999; Watanabe et al., 2001). This leaves nucleosome positioning as a potential aspect of optimal centromere organisation and indeed canonical centromeres possess highly positioned nucleosomes while neocentromeres do not. The central domain regions at fission yeast centromeres are mostly non-repetitive, in contrast to repetitive centromeres in higher eukaryotes, but the spaced non-repetitive nucleosome positioning may be functionally similar to nucleosome phasing by short centromere repeat elements (Dawe and Henikoff, 2006). *S. pombe* and *S. octosporus* central domains display highly positioned nucleosomes, but their DNA sequence is not conserved. Thus these sequences themselves might not be actually necessary to form a functional centromere. However, these sequences may still conserve ‘hidden’ features that drive processes such as transcription and factor binding that promote CENP-A chromatin/kinetochore assembly (Catania et al., 2015). Additionally, the results presented suggest that providing binding sites for ORC may also be an important aspect of canonical centromere organisation, though the actual function of ORC within the central domain remains unknown.

What does this mean for neocentromere evolution in fission yeast? Conceptually it would be of great interest to follow a neocentromere during its formation and maturation to determine whether it evolves towards an organisation similar to canonical centromere with highly positioned nucleosomes. Of course, this assumes that centromere evolution would occur gradually rather than due to sequence duplication that replaces or inactivates the neocentromere. There is only one described fission yeast isolate with four chromosomes, while all other isolates have three chromosomes (Brown et al., 2011, 2014). The ENC on the new chromosome is a duplication of the centromere of chromosome II and is not associated with heterochromatin, though the canonical and duplicated centromeres share ~100 kb of flanking non-repetitive sequences of each side (Brown et al., 2011, 2014). However, it was proposed that this did not represent a duplication that replaced a neocentromere, but rather this arose from recombination of a chromosomal rearrangement that created a dicentric chromosome and was resolved by chromosome breakage. However, it remains possible that neocentromere sequences would be replaced by duplication of a canonical centromere over evolutionary timescales.

6.2. Replication at centromeres

ORC localises to the fission yeast central domains with which CENP-A^{Cnp1} is associated (Hayashi et al., 2007; Segurado et al., 2003). The early replication of centromeres in fission yeast is due to the firing of origins of replication located within the outer-repeats and limited (if any) origin activity was detected within the central domains of endogenous centromeres (Kim et al., 2003; Smith et al., 1995). Central domain sequences are able to act as replication origins on ectopic plasmids but do not initiate replication at their endogenous loci (Smith et al., 1995; Takahashi et al., 1992). It is possible that the high AT-content of these sequences is adequate to recruit ORC and it was initially proposed that central domain sequences do not initiate replication at their endogenous loci because ORC is excluded from binding to these regions and that these sequences are important for centromere related functions (Smith et al., 1995). Of course, the observation that ORC is in fact enriched within the central domains completely recontextualises the question of why these sequences do not initiate DNA replication at their endogenous loci (Hayashi et al., 2007; Segurado et al., 2003). This opens up the possibility that ORC performs a

function at centromeres that is not related to initiating DNA replication. None of the proteins tested in this thesis were found to localise specifically to the gaps between CENP-A^{Cnp1}, but the asymmetric AT-rich sequences are the preferred binding site of ORC and it seems plausible that ORC localises to these gaps. But what might the function of ORC be within these regions?

Apart from its role in replication, ORC has been suggested to play a role in heterochromatin structure and transcriptional silencing in various systems. In humans and *Drosophila* ORC and HP1 co-localize and bind interdependently in heterochromatic regions (Auth et al., 2006; Badugu et al., 2005; Pak et al., 1997; Prasanth et al., 2010; Shareef et al., 2001, 2003). Additionally, in *Drosophila* ORC and HOAP function at telomeres to maintain telomere structure and prevent chromosome fusions and ring formation (Cenci et al., 2003). Thus, metazoan ORC may act as platform for recruiting heterochromatin components similar to its interaction with the silent information regulator (Sir) proteins in *S. cerevisiae* (Shareef et al., 2003). *S. cerevisiae* lacks H3K9 methylation and many heterochromatic proteins, with no apparent orthologs for Suv39, HP1, Dicer, or Argonaute (Harrison et al., 2009). Instead Sir1-4 are responsible for silencing at the cryptic mating type loci and telomeres (Hickman et al., 2011). At the repressed mating type loci (*HML* and *HMR*) Sir1 interacts with Orc1 to anchor the other Sir proteins (Hou et al., 2005; Zhang et al., 2002). It is possible that in fission yeast ORC plays a structural role as a platform for recruiting other proteins and activities to the central domain. Indeed, it is even possible that ORC acts as a platform for stabilising interactions between CENP-A and CENP-C, similar to the role of CENP-B in mammals (Fachinetti et al., 2015).

Fission yeast ORC has been shown to associate with the heterochromatic outer-repeats and is responsible for the early replication of centromeres (Li et al., 2011). It remains possible that the ORC pool located within the central domain forms dormant replication origins that could fire in the absence of outer-repeat origin activity to ensure proper replication timing of the centromeres. This is consistent with the observation that central domain sequences can initiate DNA replication on ectopic plasmids (Smith et al., 1995; Takahashi et al., 1992). Moreover, centromere replication timing is distinct, suggesting that control of centromere replication timing may also be of functional importance. Centromeres replicate early during S-phase in

budding yeast (Kitamura et al., 2007), fission yeast (Hayashi et al., 2009), and *C. albicans* (Koren et al., 2010). Neocentromere regions in *C. albicans* (Koren et al., 2010) change their replication timing to match the early replication of canonical centromeres. Similarly ectopic centromeres in budding yeast alter the replication timing of surrounding chromatin from late to early (Natsume et al., 2013; Pohl et al., 2012). In contrast, centromere replication takes place in mid to late S-phase in *Drosophila* (Sullivan and Karpen, 2001), chicken (Shang et al., 2013) and human cells (Ten Hagen et al., 1990; Hultdin, 2001). Chicken neocentromere formation is associated with a shift towards later replication (Shang et al., 2013). The functional significance of centromere replication timing is still not understood. There is also conflicting findings with respect to whether the ORC located within the fission yeast central domain recruits other replication proteins and forms licensed replication origins (Hayashi et al., 2007; Segurado et al., 2003). Transcription has been shown to disrupt licensed replication origins and replication proteins must be re-loaded and origins re-licensed (Lõoke et al., 2010). Pervasive low quality transcription associated with stalled RNAPII has been detected within the central domain of S. pombe central domains (Catania et al., 2015; Choi et al., 2011) and this low level central domain transcription could potentially allow ORC to bind DNA but prevent subsequent origin licensing. Alternatively, stalled RNAPII is required for ORC binding to origins within the rDNA repeats in fission yeast so transcription could instead be required for ORC binding (Mayan, 2013). Most simply, ORC could form dormant licensed origins that do not fire simply due to their proximity to early firing origins in the outer-repeats. However, the high enrichment of asymmetric AT-rich sites that could act as a binding site for the Orc4 AT-hook and enrichment of ORC across the entire central domains seems excessive simply for the formation of dormant replication origins. Alternately, since RNAPII stalling may be associated with remodelling events that promote CENP-A^{Cnp1} incorporation, it is possible that ORC could contribute to transcriptional stalling and therefore to CENP-A^{Cnp1} incorporation.

Given that the sequence features influencing nucleosome positioning are conserved between *S. pombe* and *S. octosporus*, it is also possible that ORC contributes to CENP-A^{Cnp1} nucleosome positioning. Asymmetric AT-rich sequences, such as those found within the central domains of fission yeast centromere, are preferred binding sites for ORC (Chuang and Kelly, 1999), but they are also strongly predicted to

exclude nucleosomes (Struhl and Segal, 2013). This raises the question of whether these sequences are present within centromeres to allow ORC to bind or whether they have a role in CENP-A^{Cnp1} positioning. Indeed, the AT-richness of replication origins themselves may have arisen to exclude nucleosomes from origins even in the absence of ORC and ORC recruitment itself may simply be tied to such nucleosome depleted regions (Xu et al., 2012). Metazoan replication origins are not AT-rich, but instead are enriched for a G-rich motif, termed origin G-rich repeated elements (Cayrou et al., 2012). These may nevertheless perform a similar nucleosome excluding function since poly-G tracks also known to exclude nucleosomes in *Ascomycota* species (Tsankov et al., 2011). In mice high GC-content sequences exclude nucleosomes both *in vivo* and *in vitro* (Fenouil et al., 2012). However, *in vitro* nucleosome positioning around nucleosome depleted regions at promoters does not reflect the high positioning observed *in vivo* and thus sequence alone is likely insufficient for positioning of nucleosomes (Mavrich et al., 2008). Thus, ORC may position flanking CENP-A^{Cnp1} nucleosomes within the central domain in a process similar to that observed at *S. cerevisiae* origins (Lipford and Bell, 2001). Indeed, centromeric nucleosome positioning may in fact be functionally similar to the potential role of CENP-B in nucleosome positioning at mammalian centromeres (Ando et al., 2002; Henikoff et al., 2015; Yoda et al., 1998). A comparison between *in vivo* and *in vitro* CENP-A^{Cnp1} nucleosome positioning would determine whether sequence features alone are able to position CENP-A^{Cnp1} nucleosomes in the presence and absence of ORC. Additionally, rapid degradation of ORC components *in vivo* could elucidate the potential role of ORC in CENP-A^{Cnp1} nucleosome positioning and maintenance as well as in centromere function.

6.3. Transcription and nucleosomes

To fully understand the roles of nucleosome positioning and ORC binding at fission yeast centromeres it is necessary to also consider the influence of transcription. Centromeric sequences have been shown to be transcribed in a number of species, including maize (Du et al., 2010), budding yeast (Ohkuni and Kitagawa, 2011), mice (Ferri et al., 2009; Kanellopoulou et al., 2005; Lehnertz et al., 2003), and humans (Chan et al., 2012; Wong et al., 2007). Centromere activity may require a low level of transcription (Scott, 2013). In budding yeast high levels of transcription impair centromere activity (Ohkuni and Kitagawa, 2011). Neocentromeres largely support a connection between low levels of transcription and centromere activity. Chicken

neocentromeres preferentially form at inactive genes and are associated with transcriptional inactivation of active genes following neocentromere formation (Shang et al., 2013). *C. albicans* centromere distal neocentromeres preferentially form in large intergenic regions (Ketel et al., 2009) and centromere proximal neocentromeres preferentially form at transcriptionally inactive regions (Thakur and Sanyal, 2013). Fission yeast neocentromeres form over genes that are transcribed at low levels (Ishii et al., 2008).

Transcription of fission yeast outer-repeats is clearly required for RNAi-mediated heterochromatin formation (Creamer and Partridge; Volpe et al., 2002), but transcription of the central domain of fission yeast centromeres is not as well understood. Transcription itself may be important for centromere function or alternatively promoters within the central domain that recruit the chromatin remodeller $Chd1^{Hrp1}$ may initiate transcription as a by-product of H3 eviction and CENP-A^{Cnp1} incorporation (Choi et al., 2011). The organisation of central domain sequences also contributes to transcription since it was found that transcriptional stalling at AT-rich sequences within the central domain may influence CENP-A^{Cnp1} establishment and incorporation (Catania et al., 2015). This suggests that the AT-rich sequences in the central domain could play a role in promoting transcriptional stalling, in addition to potential roles in nucleosome positioning and ORC binding.

Again, these processes are not separate since transcription itself also contributes to nucleosome positioning. Only a subset of *in vivo* budding yeast nucleosome depleted regions at promoters are also nucleosome depleted *in vitro* (Zhang et al., 2009). Additionally, the +1 nucleosome is not highly positioned *in vitro* and positioning *in vivo* is relative to the transcriptional start site and not the boundary of the nucleosome depleted region (Zhang et al., 2009). Transcription is required for proper nucleosome positioning downstream of the transcription start site and the size of the nucleosome depleted region (Weiner et al., 2010). Positioning of the +1 nucleosome on a given sequence is species specific and not simply sequence dependent (Hughes et al., 2012). In human cells stalled RNAPII at transcriptionally inactive genes was shown to be also associated with positioned nucleosomes, though the positioning differed from nucleosome positioning at transcriptionally active genes (Schones et al., 2008). Thus, it is possible that transcription could

contribute to the gaps between CENP-A^{Cnp1} nucleosomes or positioning of CENP-A^{Cnp1} nucleosomes.

6.4. Outlook

Centromere identity and function exists at the intersection of a number of biological processes, including transcription, nucleosome positioning, DNA replication, and CENP-A^{Cnp1} nucleosome incorporation. Determining how these processes individually and collectively relate to overall centromere identity and function will continue to be a challenging undertaking. Are AT-rich sequences important for recruiting ORC, positioning nucleosomes, stalling transcription, or all three? And what is the relationship between these three processes? Is transcription necessary to prevent origin licensing or for H3 nucleosome eviction and CENP-A^{Cnp1} incorporation and positioning? Does ORC position flanking nucleosomes, ensure centromere replication timing, or function in chromatin or even kinetochore structure? Is CENP-A^{Cnp1} nucleosome positioning with gaps important for allowing other complexes to bind between CENP-A^{Cnp1} nucleosomes or in mediating interactions with the kinetochore? Questions such as these are difficult to address in organisms that possess centromeres composed of repetitive sequences. Fission yeast non-repetitive central domains are deceptively simple since they are more tractable and yet they are incredibly information dense. However, it is possible to begin untangling the processes that contribute to centromere organisation using minichromosomes containing synthetic centromeres that can be engineered to have sequence features that separate these processes. In addition, it is now possible to synthesise large domains such as an entire 8 kb central domain. Thus versions with ‘gaps’ lacking features such as asymmetric AT-rich motifs could be used to test the role of such motifs. The analyses presented in this thesis (and subsequent investigation of the questions raised herein) will aid our understanding of the processes that contribute to centromere identity and indeed to epigenetic mechanisms in general.

Acknowledgements

I would like to thank Robin for all the support that went into the work presented here (and lots of work that is not). I would also like to thank Harald for teaching me how to do lab work when I was absolutely helpless. I would like to thank Robin and Alison for all their help in preparing this thesis. Thanks to all of the current and former members of the Allshire lab for lots of support over the years and to all the other scientists I directly or indirectly learned from during that time. And, of course, to my Buffalos family. And to my actual family.

References

Abendroth, C., Hofmeister, A., Hake, S.B., Kamweru, P.K., Miess, E., Dornblut, C., Küffner, I., Deng, W., Leonhardt, H., Orthaus, S., et al. (2015). The CENP-T C-Terminus Is Exclusively Proximal to H3.1 and not to H3.2 or H3.3. Int. J. Mol. Sci. *16*, 5839–5863.

Akiyoshi, B., and Gull, K. (2013). Evolutionary cell biology of chromosome segregation: insights from trypanosomes. Open Biol. *3*, 130023.

Akiyoshi, B., and Gull, K. (2014). Discovery of unconventional kinetochores in kinetoplastids. Cell *156*, 1247–1258.

Aladjem, M.I. (2007). Replication in context: dynamic regulation of DNA replication patterns in metazoans. Nat. Rev. Genet. *8*, 588–600.

Allshire, R.C., and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat. Rev. Genet. *9*, 923–937.

Allshire, R.C., Javerzat, J.P., Redhead, N.J., and Cranston, G. (1994). Position effect variegation at fission yeast centromeres. Cell *76*, 157–169.

Allshire, R.C., Nimmo, E.R., Ekwall, K., Javerzat, J.P., and Cranston, G. (1995). Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. Genes Dev. *9*, 218–233.

Alonso, A., Mahmood, R., Li, S., Cheung, F., Yoda, K., and Warburton, P.E. (2003). Genomic microarray analysis reveals distinct locations for the CENP-A binding domains in three human chromosome 13q32 neocentromeres. Hum. Mol. Genet. *12*, 2711–2721.

Alonso, A., Hasson, D., Cheung, F., and Warburton, P.E. (2010). A paucity of heterochromatin at functional human neocentromeres. Epigenetics Chromatin *3*, 6.

Ando, S., Yang, H., Nozaki, N., Okazaki, T., and Yoda, K. (2002). CENP-A, -B, and -C chromatin complex that contains the I-type alpha-satellite array constitutes the prekinetochore in HeLa cells. Mol. Cell. Biol. *22*, 2229–2241.

Appelgren, H., Kniola, B., and Ekwall, K. (2003). Distinct centromere domain structures with separate functions demonstrated in live fission yeast cells. J. Cell Sci. *116*, 4035–4042.

Aravind, L., Iyer, L.M., and Wu, C. (2007). Domain architectures of the Scm3p protein provide insights into centromere function and evolution. Cell Cycle *6*, 2511–2515.

Asano, T., Makise, M., Takehara, M., and Mizushima, T. (2007). Interaction between ORC and Cdt1p of Saccharomyces cerevisiae. FEMS Yeast Res. *7*, 1256–1262.

Asbury, C.L., Gestaut, D.R., Powers, A.F., Franck, A.D., and Davis, T.N. (2006). The Dam1 kinetochore complex harnesses microtubule dynamics to produce force and movement. Proc. Natl. Acad. Sci. U. S. A. *103*, 9873–9878.

Au, W.-C., Crisp, M.J., DeLuca, S.Z., Rando, O.J., and Basrai, M.A. (2008). Altered dosage and mislocalization of histone H3 and Cse4p lead to chromosome loss in Saccharomyces cerevisiae. Genetics *179*, 263–275.

Auth, T., Kunkel, E., and Grummt, F. (2006). Interaction between HP1alpha and replication proteins in mammalian cells. Exp. Cell Res. *312*, 3349–3359.

Badugu, R., Shareef, M.M., and Kellum, R. (2003). Novel Drosophila heterochromatin protein 1 (HP1)/origin recognition complex-associated protein (HOAP) repeat motif in HP1/HOAP interactions and chromocenter associations. J. Biol. Chem. *278*, 34491–34498.

Badugu, R., Yoo, Y., Singh, P.B., and Kellum, R. (2005). Mutations in the heterochromatin protein 1 (HP1) hinge domain affect HP1 protein interactions and chromosomal distribution. Chromosoma *113*, 370–384.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. *2*, 28–36.

Baker, R.E., and Rogers, K. (2005). Genetic and genomic analysis of the AT-rich centromere DNA element II of Saccharomyces cerevisiae. Genetics *171*, 1463–1475.

Barnhart, M.C., Kuich, P.H.J.L., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E., and Foltz, D.R. (2011). HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. J. Cell Biol. *194*, 229–243.

Barry, A. (1999). Sequence analysis of an 80 kb human neocentromere. Hum. Mol. Genet. *8*, 217–227.

Baum, M., Ngan, V.K., and Clarke, L. (1994). The centromeric K-type repeat and the central core are together sufficient to establish a functional Schizosaccharomyces pombe centromere. Mol. Biol. Cell *5*, 747–761.

Baum, M., Sanyal, K., Mishra, P.K., Thaler, N., and Carbon, J. (2006). Formation of functional centromeric chromatin is specified epigenetically in Candida albicans. Proc. Natl. Acad. Sci. U. S. A. *103*, 14877–14882.

Benjamini, Y., and Speed, T.P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. *40*, e72.

Bensasson, D. (2011). Evidence for a high mutation rate at rapidly evolving yeast centromeres. BMC Evol. Biol. *11*, 211.

Bensasson, D., Zarowiecki, M., Burt, A., and Koufopanou, V. (2008). Rapid evolution of yeast centromeres in the absence of drive. Genetics *178*, 2161–2167.

Bergmann, J.H., Rodríguez, M.G., Martins, N.M.C., Kimura, H., Kelly, D.A., Masumoto, H., Larionov, V., Jansen, L.E.T., and Earnshaw, W.C. (2011). Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. EMBO J. *30*, 328–340.

Bergmann, J.H., Jakubsche, J.N., Martins, N.M., Kagansky, A., Nakano, M., Kimura, H., Kelly, D.A., Turner, B.M., Masumoto, H., Larionov, V., et al. (2012). Epigenetic engineering: histone H3K9 acetylation is compatible with kinetochore structure and function. J. Cell Sci. *125*, 411–421.

Bernad, R., Sánchez, P., Rivera, T., Rodríguez-Corsino, M., Boyarchuk, E., Vassias, I., Ray-Gallet, D., Arnaoutov, A., Dasso, M., Almouzni, G., et al. (2011). Xenopus HJURP and condensin II are required for CENP-A assembly. J. Cell Biol. *192*, 569–582.

Berriman, M. (2005). The Genome of the African Trypanosome Trypanosoma brucei. Science (80-.). *309*, 416–422.

Bhattacharjee, S., Osman, F., Feeney, L., Lorenz, A., Bryer, C., and Whitby, M.C. (2013). MHF1-2/CENP-S-X performs distinct roles in centromere metabolism and genetic recombination. Open Biol. *3*, 130102.

Bjerling, P., Ekwall, K., Egel, R., and Thon, G. (2004). A novel type of silencing factor, Clr2, is necessary for transcriptional silencing at various chromosomal locations in the fission yeast Schizosaccharomyces pombe. Nucleic Acids Res. *32*, 4421–4428.

Black, B.E., and Bassett, E.A. (2008). The histone variant CENP-A and centromere specification. Curr. Opin. Cell Biol. *20*, 91–100.

Black, B.E., Foltz, D.R., Chakravarthy, S., Luger, K., Woods, V.L., and Cleveland, D.W. (2004). Structural determinants for generating centromeric chromatin. Nature *430*, 578–582.

Black, B.E., Jansen, L.E.T., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J. V., and Cleveland, D.W. (2007a). Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. Mol. Cell *25*, 309–322.

Black, B.E., Brock, M.A., Bédard, S., Woods, V.L., and Cleveland, D.W. (2007b). An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. Proc. Natl. Acad. Sci. U. S. A. *104*, 5008–5013.

Bloom, K.S., and Carbon, J. (1982). Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. Cell *29*, 305–317.

Blower, M.D., and Karpen, G.H. (2001). The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nat. Cell Biol. *3*, 730–739.

Blower, M.D., Sullivan, B.A., and Karpen, G.H. (2002). Conserved Organization of Centromeric Chromatin in Flies and Humans. Dev. Cell *2*, 319–330.

Bodor, D.L., Mata, J.F., Sergeev, M., David, A.F., Salimian, K.J., Panchenko, T., Cleveland, D.W., Black, B.E., Shah, J. V, and Jansen, L.E. (2014). The quantitative architecture of centromeric chromatin. Elife *3*, e02137.

Brogaard, K., Xi, L., Wang, J.-P., and Widom, J. (2012). A map of nucleosome positions in yeast at base-pair resolution. Nature *486*, 496–501.

Brown, W.R.A., Liti, G., Rosa, C., James, S., Roberts, I., Robert, V., Jolly, N., Tang, W., Baumann, P., Green, C., et al. (2011). A Geographically Diverse Collection of Schizosaccharomyces pombe Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity. G3 (Bethesda). *1*, 615–626.

Brown, W.R.A., Thomas, G., Lee, N.C.O., Blythe, M., Liti, G., Warringer, J., and Loose, M.W. (2014). Kinetochore assembly and heterochromatin formation occur autonomously in Schizosaccharomyces pombe. Proc. Natl. Acad. Sci. U. S. A. *111*, 1903–1908.

Bühler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873–886.

Bui, M., Dimitriadis, E.K., Hoischen, C., An, E., Quénet, D., Giebe, S., Nita-Lazar, A., Diekmann, S., and Dalal, Y. (2012). Cell-Cycle-Dependent Structural Transitions in the Human CENP-A Nucleosome In Vivo. Cell *150*, 317–326.

Burnside, R.D., Ibrahim, J., Flora, C., Schwartz, S., Tepperberg, J.H., Papenhausen, P.R., and Warburton, P.E. (2011). Interstitial deletion of proximal 8q including part of the centromere from unbalanced segregation of a paternal deletion/marker karyotype with neocentromere formation at 8p22. Cytogenet. Genome Res. *132*, 227–232.

Burrack, L.S., and Berman, J. (2012). Neocentromeres and epigenetically inherited features of centromeres. Chromosome Res.

Cadoret, J.-C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.-N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc. Natl. Acad. Sci. U. S. A. *105*, 15837–15842.

Camahort, R., Li, B., Florens, L., Swanson, S.K., Washburn, M.P., and Gerton, J.L. (2007). Scm3 is essential to recruit the histone h3 variant cse4 to centromeres and to maintain a functional kinetochore. Mol. Cell *26*, 853–865.

Camahort, R., Shivaraju, M., Mattingly, M., Li, B., Nakanishi, S., Zhu, D., Shilatifard, A., Workman, J.L., and Gerton, J.L. (2009). Cse4 is part of an octameric nucleosome in budding yeast. Mol. Cell *35*, 794–805.

Capozzi, O., Purgato, S., D'Addabbo, P., Archidiacono, N., Battaglia, P., Baroncini, A., Capucci, A., Stanyon, R., Della Valle, G., and Rocchi, M. (2009). Evolutionary descent of a human chromosome 6 neocentromere: a jump back to 17 million years ago. Genome Res. *19*, 778–784.

Carbone, L., Nergadze, S.G., Magnani, E., Misceo, D., Francesca Cardone, M., Roberto, R., Bertoni, L., Attolini, C., Francesca Piras, M., de Jong, P., et al. (2006). Evolutionary movement of centromeres in horse, donkey, and zebra. Genomics *87*, 777–782.

Carroll, C.W., Milks, K.J., and Straight, A.F. (2010). Dual recognition of CENP-A nucleosomes is required for centromere assembly. J. Cell Biol. *189*, 1143–1155.

Casola, C., Hucks, D., and Feschotte, C. (2008). Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. Mol. Biol. Evol. *25*, 29–41.

Castillo, A.G., Pidoux, A.L., Catania, S., Durand-Dubief, M., Choi, E.S., Hamilton, G., Ekwall, K., and Allshire, R.C. (2013). Telomeric repeats facilitate CENP-A(Cnp1) incorporation via telomere binding proteins. PLoS One *8*, e69673.

Catania, S., Pidoux, A.L., and Allshire, R.C. (2015). Sequence Features and Transcriptional Stalling within Centromere DNA Promote Establishment of CENP-A Chromatin. PLoS Genet. *11*, e1004986.

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., et al. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res. *21*, 1438–1449.

Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., and Méchali, M. (2012). New insights into replication origin characteristics in metazoans. Cell Cycle *11*, 658–667.

Cenci, G., Siriaco, G., Raffa, G.D., Kellum, R., and Gatti, M. (2003). The Drosophila HOAP protein is required for telomere capping. Nat. Cell Biol. *5*, 82–84.

Chan, F.L., Marshall, O.J., Saffery, R., Won Kim, B., Earle, E., Choo, K.H.A., and Wong, L.H. (2012). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc. Natl. Acad. Sci. U. S. A. *109*, 1979–1984.

Chang, F., May, C.D., Hoggard, T., Miller, J., Fox, C.A., and Weinreich, M. (2011). High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. Nucleic Acids Res. *39*, 6523–6535.

Cheeseman, I.M., Enquist-Newman, M., Müller-Reichert, T., Drubin, D.G., and Barnes, G. (2001a). Mitotic spindle integrity and kinetochore function linked by the Duo1p/Dam1p complex. J. Cell Biol. *152*, 197–212.

Cheeseman, I.M., Brew, C., Wolyniak, M., Desai, A., Anderson, S., Muster, N., Yates, J.R., Huffaker, T.C., Drubin, D.G., and Barnes, G. (2001b). Implication of a novel multiprotein Dam1p complex in outer kinetochore function. J. Cell Biol. *155*, 1137–1145.

Cheeseman, I.M., Chappie, J.S., Wilson-Kubalek, E.M., and Desai, A. (2006). The conserved KMN network constitutes the core microtubule-binding site of the kinetochore. Cell *127*, 983–997.

Chen, C.-C., Dechassa, M.L., Bettini, E., Ledoux, M.B., Belisario, C., Heun, P., Luger, K., and Mellone, B.G. (2014). CAL1 is the Drosophila CENP-A assembly factor. J. Cell Biol. *204*, 313–329.

Chen, E.S., Zhang, K., Nicolas, E., Cam, H.P., Zofall, M., and Grewal, S.I.S. (2008). Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature *451*, 734–737.

Chen, S., de Vries, M.A., and Bell, S.P. (2007). Orc6 is required for dynamic recruitment of Cdt1 during repeated Mcm2-7 loading. Genes Dev. *21*, 2897–2907.

Chen, Y., Baker, R.E., Keith, K.C., Harris, K., Stoler, S., and Fitzgerald-Hayes, M. (2000). The N Terminus of the Centromere H3-Like Protein Cse4p Performs an Essential Function Distinct from That of the Histone Fold Domain. Mol. Cell. Biol. *20*, 7037–7048.

Chmátal, L., Gabriel, S.I., Mitsainas, G.P., Martínez-Vargas, J., Ventura, J., Searle, J.B., Schultz, R.M., and Lampson, M.A. (2014). Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. Curr. Biol. *24*, 2295–2300.

Choi, E.S., Strålfors, A., Castillo, A.G., Durand-Dubief, M., Ekwall, K., and Allshire, R.C. (2011). Identification of noncoding transcripts from within CENP-A chromatin at fission yeast centromeres. J. Biol. Chem. *286*, 23600–23607.

Choi, E.S., Strålfors, A., Catania, S., Castillo, A.G., Svensson, J.P., Pidoux, A.L., Ekwall, K., and Allshire, R.C. (2012). Factors That Promote H3 Chromatin Integrity during Transcription Prevent Promiscuous Deposition of CENP-ACnp1 in Fission Yeast. PLoS Genet. *8*, e1002985.

Choo, K.H., Vissel, B., Nagy, A., Earle, E., and Kalitsis, P. (1991). A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. Nucleic Acids Res. *19*, 1179–1182.

Chuang, R.Y., and Kelly, T.J. (1999). The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. Proc. Natl. Acad. Sci. U. S. A. *96*, 2656–2661.

Chueh, A.C., Wong, L.H., Wong, N., and Choo, K.H.A. (2005). Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. Hum. Mol. Genet. *14*, 85–93.

Chueh, A.C., Northrop, E.L., Brettingham-Moore, K.H., Choo, K.H.A., and Wong, L.H. (2009). LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin. PLoS Genet. *5*, e1000354.

Chung, H.-R., Dunkel, I., Heise, F., Linke, C., Krobitsch, S., Ehrenhofer-Murray, A.E., Sperling, S.R., and Vingron, M. (2010). The effect of micrococcal nuclease digestion on nucleosome positioning data. PLoS One *5*, e15754.

Ciferri, C., Pasqualato, S., Screpanti, E., Varetti, G., Santaguida, S., Dos Reis, G., Maiolica, A., Polka, J., De Luca, J.G., De Wulf, P., et al. (2008). Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. Cell *133*, 427–439.

Clarke, L., and Baum, M.P. (1990). Functional analysis of a centromere from fission yeast: a role for centromere-specific repeated DNA sequences. Mol. Cell. Biol. *10*, 1863–1872.

Clarke, L., and Carbon, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature *287*, 504–509.

Clarke, L., Amstutz, H., Fishel, B., and Carbon, J. (1986). Analysis of centromeric DNA in the fission yeast Schizosaccharomyces pombe. Proc. Natl. Acad. Sci. U. S. A. *83*, 8253–8257.

Coffman, V.C., Wu, P., Parthun, M.R., and Wu, J.-Q. (2011). CENP-A exceeds microtubule attachment sites in centromere clusters of both budding and fission yeast. J. Cell Biol. *195*, 563–572.

Cohen, R.L., Espelin, C.W., De Wulf, P., Sorger, P.K., Harrison, S.C., and Simons, K.T. (2008). Structural and functional dissection of Mif2p, a conserved DNA-binding kinetochore protein. Mol. Biol. Cell *19*, 4480–4491.

Comings, D.E., and Okada, T.A. (1971). Fine structure of kinetochore in Indian muntjac. Exp. Cell Res. *67*, 97–110.

Conde e Silva, N., Black, B.E., Sivolob, A., Filipski, J., Cleveland, D.W., and Prunell, A. (2007). CENP-A-containing nucleosomes: easier disassembly versus exclusive centromeric localization. J. Mol. Biol. *370*, 555–573.

Cooke, C.A. (1993). Mapping DNA within the mammalian kinetochore. J. Cell Biol. *120*, 1083–1091.

Cottarel, G., Shero, J.H., Hieter, P., and Hegemann, J.H. (1989). A 125-base-pair CEN6 DNA fragment is sufficient for complete meiotic and mitotic centromere functions in Saccharomyces cerevisiae. Mol. Cell. Biol. *9*, 3342–3349.

Courbet, S., Gay, S., Arnoult, N., Wronka, G., Anglana, M., Brison, O., and Debatisse, M. (2008). Replication fork movement sets chromatin loop size and origin choice in mammalian cells. Nature *455*, 557–560.

Creamer, K.M., and Partridge, J.F. RITS-connecting transcription, RNA interference, and heterochromatin assembly in fission yeast. Wiley Interdiscip. Rev. RNA *2*, 632–646.

Dalal, Y., Wang, H., Lindsay, S., and Henikoff, S. (2007). Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. PLoS Biol. *5*, e218.

Dambacher, S., Deng, W., Hahn, M., Sadic, D., Fröhlich, J.J., Nuber, A., Hoischen, C., Diekmann, S., Leonhardt, H., and Schotta, G. (2012). CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin Do not distribute . © 2012 Landes Bioscience . 101–110.

Dawe, R.K., and Henikoff, S. (2006). Centromeres put epigenetics in the driver's seat. Trends Biochem. Sci. *31*, 662–669.

Debeauchamp, J.L., Moses, A., Noffsinger, V.J.P., Ulrich, D.L., Job, G., Kosinski, A.M., and Partridge, J.F. (2008). Chp1-Tas3 interaction is required to recruit RITS to fission yeast centromeres and for maintenance of centromeric heterochromatin. Mol. Cell. Biol. *28*, 2154–2166.

Delgado, S., Gómez, M., Bird, A., and Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J. *17*, 2426–2435.

Dhar, S.K., and Dutta, A. (2000). Identification and characterization of the human ORC6 homolog. J. Biol. Chem. *275*, 34983–34988.

Dimitriadis, E.K., Weber, C., Gill, R.K., Diekmann, S., and Dalal, Y. (2010). Tetrameric organization of vertebrate centromeric nucleosomes. Proc. Natl. Acad. Sci. U. S. A. *107*, 20317–20322.

Ding, R., McDonald, K.L., and McIntosh, J.R. (1993). Three-dimensional reconstruction and analysis of mitotic spindles from the yeast, Schizosaccharomyces pombe. J. Cell Biol. *120*, 141–151.

Dingwall, C., Lomonossoff, G.P., and Laskey, R.A. (1981). High sequence specificity of micrococcal nuclease. Nucleic Acids Res. *9*, 2659–2673.

Djupedal, I., Portoso, M., Spåhr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. (2005). RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. Genes Dev. *19*, 2301–2306.

Djupedal, I., Kos-Braun, I.C., Mosher, R. a, Söderholm, N., Simmer, F., Hardcastle, T.J., Fender, A., Heidrich, N., Kagansky, A., Bayne, E., et al. (2009). Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. EMBO J. *28*, 3832–3844.

Dong, Y., Vanden Beldt, K.J., Meng, X., Khodjakov, A., and McEwen, B.F. (2007). The outer plate in vertebrate kinetochores is a flexible network with multiple microtubule interactions. Nat. Cell Biol. *9*, 516–522.

Donovan, S., Harwood, J., Drury, L.S., and Diffley, J.F.X. (1997). Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. Proc. Natl. Acad. Sci. U. S. A. *94*, 5611–5616.

Dornblut, C., Quinn, N., Monajambashi, S., Prendergast, L., van Vuuren, C., Münch, S., Deng, W., Leonhardt, H., Cardoso, M.C., Hoischen, C., et al. (2014). A CENP-S/X complex assembles at the centromere in S and G2 phases of the human cell cycle. Open Biol. *4*, 130229.

Drinnenberg, I.A., deYoung, D., Henikoff, S., and Malik, H.S. (2014). Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. Elife *3*.

Du, Y., Topp, C.N., and Dawe, R.K. (2010). DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. PLoS Genet. *6*, e1000835.

Duncker, B.P., Chesnokov, I.N., and McConkey, B.J. (2009). The origin recognition complex protein family. Genome Biol. *10*, 214.

Dunleavy, E.M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., Daigo, Y., Nakatani, Y., and Almouzni-Pettinotti, G. (2009). HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. Cell *137*, 485–497.

Dunleavy, E.M., Almouzni, G., and Karpen, G.H. (2011). H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G1 phase. Nucleus *2*, 146–157.

Dunleavy, E.M., Zhang, W., and Karpen, G.H. (2013). Solo or doppio: how many CENP-As make a centromeric nucleosome? Nat. Struct. Mol. Biol. *20*, 648–650.

Dutrillaux, B. (1979). Chromosomal evolution in primates: tentative phylogeny from Microcebus murinus (Prosimian) to man. Hum. Genet. *48*, 251–314.

Earnshaw, W.C., and Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. Chromosoma *91*, 313–321.

Earnshaw, W.C., Ratrie, H., and Stetten, G. (1989). Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. Chromosoma *98*, 1–12.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. Genes Dev. *24*, 748–753.

Eaton, M.L., Prinz, J.A., MacAlpine, H.K., Tretyakov, G., Kharchenko, P. V, and MacAlpine, D.M. (2011). Chromatin signatures of the Drosophila replication program. Genome Res. *21*, 164–174.

Echeverry, M.C., Bot, C., Obado, S.O., Taylor, M.C., and Kelly, J.M. (2012). Centromere-associated repeat arrays on Trypanosoma brucei chromosomes are much more extensive than predicted. BMC Genomics *13*, 29.

Edwards, M.C., Tutter, A. V, Cvetic, C., Gilbert, C.H., Prokhorova, T.A., and Walter, J.C. (2002a). MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in Xenopus egg extracts. J. Biol. Chem. *277*, 33049–33057.

Edwards, M.C., Tutter, A. V, Cvetic, C., Gilbert, C.H., Prokhorova, T.A., and Walter, J.C. (2002b). MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in Xenopus egg extracts. J. Biol. Chem. *277*, 33049–33057.

Ekwall, K., Nimmo, E., Javerzat, J., Borgstrom, B., Egel, R., Cranston, G., and Allshire, R. (1996). Mutations in the fission yeast silencing factors clr4+ and rik1+ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. J. Cell Sci. *109*, 2637–2648.

Euskirchen, G.M. (2002). Nnf1p, Dsn1p, Mtw1p, and Nsl1p: a new group of proteins important for chromosome segregation in Saccharomyces cerevisiae. Eukaryot. Cell *1*, 229–240.

Fachinetti, D., Diego Folco, H., Nechemia-Arbely, Y., Valente, L.P., Nguyen, K., Wong, A.J., Zhu, Q., Holland, A.J., Desai, A., Jansen, L.E.T., et al. (2013). A two-step mechanism for epigenetic specification of centromere identity and function. Nat. Cell Biol.

Fachinetti, D., Han, J.S., McMahon, M.A., Ly, P., Abdullah, A., Wong, A.J., and Cleveland, D.W. (2015). DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. Dev. Cell *33*, 314–327.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I., et al. (2012). CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res. *22*, 2399–2408.

Ferri, F., Bouzinba-Segard, H., Velasco, G., Hubé, F., and Francastel, C. (2009). Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. Nucleic Acids Res. *37*, 5071–5080.

Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I.K., Sharon, E., Lubling, Y., Widom, J., and Segal, E. (2008). Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Comput. Biol. *4*, e1000216.

Fishel, B., Amstutz, H., Baum, M., Carbon, J., and Clarke, L. (1988). Structural organization and functional analysis of centromeric DNA in the fission yeast Schizosaccharomyces pombe. Mol. Cell. Biol. *8*, 754–763.

Fitzgerald-Hayes, M., Clarke, L., and Carbon, J. (1982). Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. Cell *29*, 235–244.

Folco, H.D., Pidoux, A.L., Urano, T., and Allshire, R.C. (2008). Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. Science *319*, 94–97.

Foltz, D.R., Jansen, L.E.T., Black, B.E., Bailey, A.O., Yates, J.R., and Cleveland, D.W. (2006a). The human CENP-A centromeric nucleosome-associated complex. Nat. Cell Biol. *8*, 458–469.

Foltz, D.R., Jansen, L.E.T., Black, B.E., Bailey, A.O., Yates, J.R., and Cleveland, D.W. (2006b). The human CENP-A centromeric nucleosome-associated complex. Nat. Cell Biol. *8*, 458–469.

Foltz, D.R., Jansen, L.E.T., Bailey, A.O., Yates, J.R., Bassett, E.A., Wood, S., Black, B.E., and Cleveland, D.W. (2009). Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. Cell *137*, 472–484.

Forsburg, S.L. (1993). Comparison of Schizosaccharomyces pombe expression systems. Nucleic Acids Res. *21*, 2955–2956.

Foss, M., McNally, F.J., Laurenson, P., and Rine, J. (1993). Origin recognition complex (ORC) in transcriptional silencing and DNA replication in S. cerevisiae. Science *262*, 1838–1844.

Fujita, Y., Hayashi, T., Kiyomitsu, T., Toyoda, Y., Kokubu, A., Obuse, C., and Yanagida, M. (2007). Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. Dev. Cell *12*, 17–30.

Fukagawa, T., and Brown, W.R. (1997). Efficient conditional mutation of the vertebrate CENP-C gene. Hum. Mol. Genet. *6*, 2301–2308.

Fukagawa, T., and Earnshaw, W.C. (2014). The Centromere: Chromatin Foundation for the Kinetochore Machinery. Dev. Cell *30*, 496–508.

Fukagawa, T., Pendon, C., Morris, J., and Brown, W. (1999). CENP-C is necessary but not sufficient to induce formation of a functional centromere. EMBO J. *18*, 4196–4209.

Fukagawa, T., Mikami, Y., Nishihashi, A., Regnier, V., Haraguchi, T., Hiraoka, Y., Sugata, N., Todokoro, K., Brown, W., and Ikemura, T. (2001). CENP-H, a constitutive centromere component, is required for centromere targeting of CENP-C in vertebrate cells. EMBO J. *20*, 4603–4617.

Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., and Oshimura, M. (2004). Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nat. Cell Biol. *6*, 784–791.

Furuyama, S., and Biggins, S. (2007). Centromere identity is specified by a single centromeric nucleosome in budding yeast. Proc. Natl. Acad. Sci. U. S. A. *104*, 14706–14711.

Gaczynska, M., Osmulski, P.A., Jiang, Y., Lee, J.-K., Bermudez, V., and Hurwitz, J. (2004). Atomic force microscopic analysis of the binding of the Schizosaccharomyces pombe origin recognition complex and the spOrc4 protein with origin DNA. Proc. Natl. Acad. Sci. U. S. A. *101*, 17952–17957.

Gaitanos, T.N., Santamaria, A., Jeyaprakash, A.A., Wang, B., Conti, E., and Nigg, E.A. (2009). Stable kinetochore-microtubule interactions depend on the Ska complex and its new component Ska3/C13Orf3. EMBO J. *28*, 1442–1452.

Gallagher, J.E.G., Babiarz, J.E., Teytelman, L., Wolfe, K.H., and Rine, J. (2009). Elaboration, diversification and regulation of the Sir1 family of silencing proteins in Saccharomyces. Genetics *181*, 1477–1491.

Gardner, K.A., Rine, J., and Fox, C.A. (1999). A region of the Sir1 protein dedicated to recognition of a silencer and required for interaction with the Orc1 protein in saccharomyces cerevisiae. Genetics *151*, 31–44.

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T., and Cheeseman, I.M. (2011). Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. Cell *145*, 410–422.

Gasser, S.M., and Cockell, M.M. (2001). The molecular biology of the SIR proteins. Gene *279*, 1–16.

Gassmann, R., Rechtsteiner, A., Yuen, K.W., Muroyama, A., Egelhofer, T., Gaydos, L., Barron, F., Maddox, P., Essex, A., Monen, J., et al. (2012). An inverse relationship to germline transcription defines centromeric chromatin in C. elegans. Nature *484*, 534–537.

Ge, X.Q., Jackson, D.A., and Blow, J.J. (2007). Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. Genes Dev. *21*, 3331–3341.

Giri, S., Aggarwal, V., Pontis, J., Shen, Z., Chakraborty, A., Khan, A., Mizzen, C., Prasanth, K. V, Ait-Si-Ali, S., Ha, T., et al. (2015). The preRC protein ORCA organizes heterochromatin by assembling histone H3 lysine 9 methyltransferases on chromatin. Elife *4*.

Goldberg, I.G., Sawhney, H., Pluta, A.F., Warburton, P.E., and Earnshaw, W.C. (1996). Surprising deficiency of CENP-B binding sites in African green monkey alpha-satellite DNA: implications for CENP-B function at centromeres. Mol. Cell. Biol. *16*, 5156–5168.

Gómez, M., and Antequera, F. (2008). Overreplication of short DNA regions during S phase in human cells. Genes Dev. *22*, 375–385.

Grallert, B., and Nurse, P. (1996). The ORC1 homolog orp1 in fission yeast plays a key role in regulating onset of S phase. Genes Dev. *10*, 2644–2654.

Grimes, B.R., Rhoades, A.A., and Willard, H.F. (2002). Alpha-satellite DNA and vector composition influence rates of human artificial chromosome formation. Mol. Ther. *5*, 798–805.

Grimes, B.R., Babcock, J., Rudd, M.K., Chadwick, B., and Willard, H.F. (2004). Assembly and characterization of heterochromatin and euchromatin on human artificial chromosomes. Genome Biol. *5*, R89.

Haaf, T., Mater, A.G., Wienberg, J., and Ward, D.C. (1995). Presence and abundance of CENP-B box sequences in great ape subsets of primate-specific alpha-satellite DNA. J. Mol. Evol. *41*, 487–491.

Ten Hagen, K.G., Gilbert, D.M., Willard, H.F., and Cohen, S.N. (1990). Replication timing of DNA sequences associated with human centromeres and telomeres. Mol. Cell. Biol. *10*, 6348–6355.

Hahnenberger, K.M., Carbon, J., and Clarke, L. (1991). Identification of DNA regions required for mitotic and meiotic functions within the centromere of Schizosaccharomyces pombe chromosome I. Mol. Cell. Biol. *11*, 2206–2215.

Halic, M., and Moazed, D. (2010). Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. Cell *140*, 504–516.

Harris, M.E., Böhni, R., Schneiderman, M.H., Ramamurthy, L., Schümperli, D., and Marzluff, W.F. (1991). Regulation of histone mRNA in the unperturbed cell cycle: evidence suggesting control at two posttranscriptional steps. Mol. Cell. Biol. *11*, 2416–2424.

Harrison, B.R., Yazgan, O., and Krebs, J.E. (2009). Life without RNAi: noncoding RNAs and their functions in Saccharomyces cerevisiae. Biochem. Cell Biol. *87*, 767–779.

Harvey, K.J., and Newport, J. (2003). CpG Methylation of DNA Restricts Prereplication Complex Assembly in Xenopus Egg Extracts. Mol. Cell. Biol. *23*, 6769–6779.

Hasson, D., Panchenko, T., Salimian, K.J., Salman, M.U., Sekulic, N., Alonso, A., Warburton, P.E., and Black, B.E. (2013). The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat. Struct. Mol. Biol. *20*, 687–695.

Hayano, M., Kanoh, Y., Matsumoto, S., Renard-Guillet, C., Shirahige, K., and Masai, H. (2012). Rif1 is a global regulator of timing of replication origin firing in fission yeast. Genes Dev. *26*, 137–150.

Hayashi, A., Asakawa, H., Haraguchi, T., and Hiraoka, Y. (2006). Reconstruction of the kinetochore during meiosis in fission yeast Schizosaccharomyces pombe. Mol. Biol. Cell *17*, 5173–5184.

Hayashi, M., Katou, Y., Itoh, T., Tazumi, A., Tazumi, M., Yamada, Y., Takahashi, T., Nakagawa, T., Shirahige, K., and Masukata, H. (2007). Genome-wide localization of

pre-RC sites and identification of replication origins in fission yeast. EMBO J. *26*, 1327–1339.

Hayashi, M.T., Takahashi, T.S., Nakagawa, T., Nakayama, J., and Masukata, H. (2009). The heterochromatin protein Swi6/HP1 activates replication origins at the pericentromeric region and silent mating-type locus. Nat. Cell Biol. *11*, 357–362.

Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. Cell *118*, 715–729.

Hayashi, T., Ebe, M., Nagao, K., Kokubu, A., Sajiki, K., and Yanagida, M. (2014). Schizosaccharomyces pombe centromere protein Mis19 links Mis16 and Mis18 to recruit CENP-A through interacting with NMD factors and the SWI/SNF complex. Genes Cells.

Hayden, K.E., and Willard, H.F. (2012). Composition and organization of active centromere sequences in complex genomes. BMC Genomics *13*, 324.

Hegemann, J.H., Shero, J.H., Cottarel, G., Philippsen, P., and Hieter, P. (1988). Mutational analysis of centromere DNA from chromosome VI of Saccharomyces cerevisiae. Mol. Cell. Biol. *8*, 2523–2535.

Henikoff, S., and Henikoff, J.G. (2012). “Point” centromeres of Saccharomyces harbor single centromere-specific nucleosomes. Genetics *190*, 1575–1577.

Henikoff, J.G., Thakur, J., Kasinathan, S., and Henikoff, S. (2015). A unique chromatin complex occupies young α-satellite arrays of human centromeres. Sci. Adv. *1*.

Henikoff, S., Ahmad, K., and Malik, H.S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. Science *293*, 1098–1102.

Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D., and Karpen, G.H. (2006). Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. Dev. Cell *10*, 303–315.

Heus, J.J., Zonneveld, B.J.M., Steensma, H.Y., and Berg, J.A. (1990). Centromeric DNA of Kluyveromyces lactis. Curr. Genet. *18*, 517–522.

Heus, J.J., Zonneveld, B.J., de Steensma, H.Y., and van den Berg, J.A. (1993). The consensus sequence of Kluyveromyces lactis centromeres shows homology to functional centromeric DNA from Saccharomyces cerevisiae. Mol. Gen. Genet. *236*, 355–362.

Heus, J.J., Zonneveld, B.J., Steensma, H.Y., and Van den Berg, J.A. (1994). Mutational analysis of centromeric DNA elements of Kluyveromyces lactis and their role in determining the species specificity of the highly homologous centromeres from K. lactis and Saccharomyces cerevisiae. Mol. Gen. Genet. *243*, 325–333.

Hewawasam, G., Shivaraju, M., Mattingly, M., Venkatesh, S., Martin-Brown, S., Florens, L., Workman, J.L., and Gerton, J.L. (2010). Psh1 is an E3 ubiquitin ligase that targets the centromeric histone variant Cse4. Mol. Cell *40*, 444–454.

Hickman, M.A., and Rusche, L.N. (2010). Transcriptional silencing functions of the yeast protein Orc1/Sir3 subfunctionalized after gene duplication. Proc. Natl. Acad. Sci. *107*, 19384–19389.

Hickman, M. a, Froyd, C. a, and Rusche, L.N. (2011). Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage. Eukaryot. Cell *10*, 1183–1192.

Hill, A., and Bloom, K. (1987). Genetic manipulation of centromere function. Mol. Cell. Biol. *7*, 2397–2405.

Hill, A., and Bloom, K. (1989). Acquisition and processing of a conditional dicentric chromosome in Saccharomyces cerevisiae. Mol. Cell. Biol. *9*, 1368–1370.

Holoch, D., and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. Nat. Rev. Genet. *16*, 71–84.

Van Hooser, A.A., Ouspenski, I.I., Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg, M.L., Yokomori, K., Earnshaw, W.C., Sullivan, K.F., and Brinkley, B.R. (2001). Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. J. Cell Sci. *114*, 3529–3542.

Hori, T., Amano, M., Suzuki, A., Backer, C.B., Welburn, J.P., Dong, Y., McEwen, B.F., Shang, W.-H., Suzuki, E., Okawa, K., et al. (2008). CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. Cell *135*, 1039–1052.

Hörz, W., and Altenburger, W. (1981). Sequence specific cleavage of DNA by micrococcal nuclease. Nucleic Acids Res. *9*, 2643–2658.

Hou, Z., Bernstein, D.A., Fox, C.A., and Keck, J.L. (2005). Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. Proc. Natl. Acad. Sci. U. S. A. *102*, 8489–8494.

Houchens, C.R., Lu, W., Chuang, R.-Y., Frattini, M.G., Fuller, A., Simancek, P., and Kelly, T.J. (2008). Multiple mechanisms contribute to Schizosaccharomyces pombe origin recognition complex-DNA interactions. J. Biol. Chem. *283*, 30216–30224.

Hu, H., Liu, Y., Wang, M., Fang, J., Huang, H., Yang, N., Li, Y., Wang, J., Yao, X., Shi, Y., et al. (2011). Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. Genes Dev. *25*, 901–906.

Hua, X.H., and Newport, J. (1998). Identification of a preinitiation step in DNA replication that is independent of origin recognition complex and cdc6, but dependent on cdk2. J. Cell Biol. *140*, 271–281.

Hua, X.H., Yan, H., and Newport, J. (1997). A role for Cdk2 kinase in negatively regulating DNA replication during S phase of the cell cycle. J. Cell Biol. *137*, 183–192.

Hughes, A.L., and Rando, O.J. (2014). Mechanisms Underlying Nucleosome Positioning in vivo. Annu. Rev. Biophys.

Hughes, A.L.L., Jin, Y., Rando, O.J.J., and Struhl, K. (2012). A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. Mol. Cell *48*, 5–15.

Hultdin, M. (2001). Replication Timing of Human Telomeric DNA and Other Repetitive Sequences Analyzed by Fluorescence in Situ Hybridization and Flow Cytometry. Exp. Cell Res. *271*, 223–229.

Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H., and Masumoto, H. (1998). Construction of YAC-based mammalian artificial chromosomes. Nat. Biotechnol. *16*, 431–439.

Irelan, J.T., Gutkin, G.I., and Clarke, L. (2001). Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. Genetics *157*, 1191–1203.

Ishii, K., Ogiyama, Y., Chikashige, Y., Soejima, S., Masuda, F., Kakuma, T., Hiraoka, Y., and Takahashi, K. (2008). Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. Science *321*, 1088–1091.

Iyer, L.M., Leipe, D.D., Koonin, E. V, and Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. J. Struct. Biol. *146*, 11–31.

Jäger, D., and Philippsen, P. (1989). Stabilization of dicentric chromosomes in Saccharomyces cerevisiae by telomere addition to broken ends or by centromere deletion. EMBO J. *8*, 247–254.

Janke, C., Ortiz, J., Lechner, J., Shevchenko, A., Magiera, M.M., Schramm, C., and Schiebel, E. (2001). The budding yeast proteins Spc24p and Spc25p interact with Ndc80p and Nuf2p at the kinetochore and are important for kinetochore clustering and checkpoint control. EMBO J. *20*, 777–791.

Janke, C., Ortíz, J., Tanaka, T.U., Lechner, J., and Schiebel, E. (2002). Four new subunits of the Dam1-Duo1 complex reveal novel functions in sister kinetochore biorientation. EMBO J. *21*, 181–193.

Jansen, L.E.T., Black, B.E., Foltz, D.R., and Cleveland, D.W. (2007). Propagation of centromeric chromatin requires exit from mitosis. J. Cell Biol. *176*, 795–805.

Jeyaprakash, A.A., Santamaria, A., Jayachandran, U., Chan, Y.W., Benda, C., Nigg, E.A., and Conti, E. (2012). Structural and functional organization of the Ska complex, a key component of the kinetochore-microtubule interface. Mol. Cell *46*, 274–286.

Joglekar, A.P., Bouck, D., Finley, K., Liu, X., Wan, Y., Berman, J., He, X., Salmon, E.D., and Bloom, K.S. (2008). Molecular architecture of the kinetochore-microtubule attachment site is conserved between point and regional centromeres. J. Cell Biol. *181*, 587–594.

Jokelainen, P.T. (1967). The ultrastructure and spatial organization of the metaphase kinetochore in mitotic rat cells. J. Ultrastruct. Res. *19*, 19–44.

Kagansky, A., Folco, H.D., Almeida, R., Pidoux, A.L., Boukaba, A., Simmer, F., Urano, T., Hamilton, G.L., Allshire, R.C., and Allshire, C. (2009). Synthetic heterochromatin bypasses RNAi and centromeric repeats to establish functional centromeres. Science *324*, 1716–1719.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev. *19*, 489–501.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362–366.

Karpen, G.H., and Allshire, R.C. (1997). The case for epigenetic effects on centromere identity and function. Trends Genet. *13*, 489–496.

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005). RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science *309*, 467–469.

Kawasaki, Y., Kim, H.-D., Kojima, A., Seki, T., and Sugino, A. (2006). Reconstitution of Saccharomyces cerevisiae prereplicative complex assembly in vitro. Genes Cells *11*, 745–756.

Kearsey, S.E., Montgomery, S., Labib, K., and Lindner, K. (2000). Chromatin binding of the fission yeast replication factor mcm4 occurs during anaphase and requires ORC and cdc18. EMBO J. *19*, 1681–1690.

Kelley, D.E., Stokes, D.G., and Perry, R.P. (1999). CHD1 interacts with SSRP1 and depends on both its chromodomain and its ATPase/helicase-like domain for proper association with chromatin. Chromosoma *108*, 10–25.

Kellis, M., Birren, B.W., and Lander, E.S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature *428*, 617–624.

Kerres, A., Jakopec, V., and Fleig, U. (2007). The conserved Spc7 protein is required for spindle integrity and links kinetochore complexes in fission yeast. Mol. Biol. Cell *18*, 2441–2454.

Ketel, C., Wang, H.S.W., McClellan, M., Bouchonville, K., Selmecki, A., Lahav, T., Gerami-Nejad, M., and Berman, J. (2009). Neocentromeres form efficiently at multiple possible loci in Candida albicans. PLoS Genet. *5*, e1000400.

Kiermaier, E., Woehrer, S., Peng, Y., Mechtler, K., and Westermann, S. (2009). A Dam1-based artificial kinetochore is sufficient to promote chromosome segregation in budding yeast. Nat. Cell Biol. *11*, 1109–1115.

Kim, H.-S., Mukhopadhyay, R., Rothbart, S.B., Silva, A.C., Vanoosthuyse, V., Radovani, E., Kislinger, T., Roguev, A., Ryan, C.J., Xu, J., et al. (2014). Identification of a BET family bromodomain/casein kinase II/TAF-containing complex as a regulator of mitotic condensin function. Cell Rep. *6*, 892–905.

Kim, S.-M., Dubey, D.D., and Huberman, J.A. (2003). Early-replicating heterochromatin. Genes Dev. *17*, 330–335.

Kipling, D., Mitchell, A., Masumoto, H., Wilson, H., Nicol, L., and Cooke, H. (1995). CENP-B binds a novel centromeric sequence in the Asian mouse Mus caroli. Mol. Cell. Biol. *15*, 4009–4020.

Kitada, K., Yamaguchi, E., and Arisawa, M. (1996). Isolation of a Candida glabrata centromere and its use in construction of plasmid vectors. Gene *175*, 105–108.

Kitamura, E., Tanaka, K., Kitamura, Y., and Tanaka, T.U. (2007). Kinetochore microtubule interaction during S phase in Saccharomyces cerevisiae. Genes Dev. *21*, 3319–3330.

Kline, S.L., Cheeseman, I.M., Hori, T., Fukagawa, T., and Desai, A. (2006). The human Mis12 complex is required for kinetochore assembly and proper chromosome segregation. J. Cell Biol. *173*, 9–17.

Kloc, A., Zaratiegui, M., Nora, E., and Martienssen, R. (2008). RNA interference guides histone modification during the S phase of chromosomal replication. Curr. Biol. *18*, 490–495.

Kobayashi, N., Suzuki, Y., Schoenfeld, L.W., Müller, C.A., Nieduszynski, C., Wolfe, K.H., and Tanaka, T.U. (2015). Discovery of an Unconventional Centromere in Budding Yeast Redefines Evolution of Point Centromeres. Curr. Biol.

Kong, D., and DePamphilis, M.L. (2001). Site-specific DNA binding of the Schizosaccharomyces pombe origin recognition complex is determined by the Orc4 subunit. Mol. Cell. Biol. *21*, 8095–8103.

Kong, D., and DePamphilis, M.L. (2002). Site-specific ORC binding, pre-replication complex assembly and DNA synthesis at Schizosaccharomyces pombe replication origins. EMBO J. *21*, 5567–5576.

Kong, D., Coleman, T.R., and DePamphilis, M.L. (2003). Xenopus origin recognition complex (ORC) initiates DNA replication preferentially at sequences targeted by Schizosaccharomyces pombe ORC. EMBO J. *22*, 3441–3450.

Koren, A., Tsai, H.-J., Tirosh, I., Burrack, L.S., Barkai, N., and Berman, J. (2010). Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase. PLoS Genet. *6*, e1001068.

Koshland, D., Rutledge, L., Fitzgerald-Hayes, M., and Hartwell, L.H. (1987). A genetic analysis of dicentric minichromosomes in Saccharomyces cerevisiae. Cell *48*, 801–812.

Krizaic, I., Williams, S.J., Sánchez, P., Rodríguez-Corsino, M., Stukenberg, T., and Losada, A. (2015). The distinct functions of CENP-C and CENP-T/W in centromere propagation and function in Xenopus egg extracts. Nucleus.

Kuhn, R.M., Clarke, L., and Carbon, J. (1991). Clustered tRNA genes in Schizosaccharomyces pombe centromeric DNA sequence repeats. Proc. Natl. Acad. Sci. *88*, 1306–1310.

Kuipers, M.A., Stasevich, T.J., Sasaki, T., Wilson, K.A., Hazelwood, K.L., McNally, J.G., Davidson, M.W., and Gilbert, D.M. (2011). Highly stable loading of Mcm proteins onto chromatin in living cells requires replication to unload. J. Cell Biol. *192*, 29–41.

Kunkel, G.R., and Martinson, H.G. (1981). Nucleosomes will not form on double-stranded RNa or over poly(dA).poly(dT) tracts in recombinant DNA. Nucleic Acids Res. *9*, 6869–6888.

Kwon, M.-S., Hori, T., Okada, M., and Fukagawa, T. (2007). CENP-C is involved in chromosome segregation, mitotic checkpoint function, and kinetochore assembly. Mol. Biol. Cell *18*, 2155–2168.

Labib, K., Tercero, J.A., and Diffley, J.F. (2000). Uninterrupted MCM2-7 Function Required for DNA Replication Fork Progression. Science (80-.). *288*, 1643–1647.

Labib, K., Kearsey, S.E., and Diffley, J.F.X. (2001). MCM2-7 proteins are essential components of prereplicative complexes that accumulate cooperatively in the nucleus during G1-phase and are required to establish, but not maintain, the S-phase checkpoint. Mol. Biol. Cell *12*, 3658–3667.

Lacefield, S., Lau, D.T.C., and Murray, A.W. (2009). Recruiting a microtubule-binding complex to DNA directs chromosome segregation in budding yeast. Nat. Cell Biol. *11*, 1116–1120.

Lacoste, N., Woolfe, A., Tachiwana, H., Garea, A.V., Barth, T., Cantaloube, S., Kurumizaka, H., Imhof, A., and Almouzni, G. (2014). Mislocalization of the Centromeric Histone Variant CenH3/CENP-A in Human Cells Depends on the Chaperone DAXX. Mol. Cell *53*, 631–644.

Ladenburger, E.-M., Keller, C., and Knippers, R. (2002). Identification of a binding region for human origin recognition complex proteins 1 and 2 that coincides with an origin of DNA replication. Mol. Cell. Biol. *22*, 1036–1048.

Lam, A.L., Boivin, C.D., Bonney, C.F., Rudd, M.K., and Sullivan, B.A. (2006). Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. Proc. Natl. Acad. Sci. U. S. A. *103*, 4186–4191.

Lampert, F., Hornung, P., and Westermann, S. (2010). The Dam1 complex confers microtubule plus end-tracking activity to the Ndc80 kinetochore complex. J. Cell Biol. *189*, 641–649.

Lando, D., Endesfelder, U., Berger, H., Subramanian, L., Dunne, P.D., McColl, J., Klenerman, D., Carr, A.M., Sauer, M., Allshire, R.C., et al. (2012). Quantitative single-molecule microscopy reveals that CENP-ACnp1 deposition occurs during G2 in fission yeast. Open Biol. *2*, 120078–120078.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lanini, L., and McKeon, F. (1995). Domains required for CENP-C assembly at the kinetochore. Mol. Biol. Cell *6*, 1049–1059.

Lantermann, A.B., Straub, T., Strålfors, A., Yuan, G.-C., Ekwall, K., and Korber, P. (2010). Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nat. Struct. Mol. Biol. *17*, 251–257.

Lawrimore, J., Bloom, K.S., and Salmon, E.D. (2011). Point centromeres contain more than a single centromere-specific Cse4 (CENP-A) nucleosome. J. Cell Biol. *195*, 573–582.

Lee, J.K., Huberman, J.A., and Hurwitz, J. (1997). Purification and characterization of a CENP-B homologue protein that binds to the centromeric K-type repeat DNA of Schizosaccharomyces pombe. Proc. Natl. Acad. Sci. U. S. A. *94*, 8427–8432.

Lee, J.K., Moon, K.Y., Jiang, Y., and Hurwitz, J. (2001). The Schizosaccharomyces pombe origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. Proc. Natl. Acad. Sci. U. S. A. *98*, 13589–13594.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat. Genet. *39*, 1235–1244.

Lehnertz, B., Ueda, Y., Derijck, A.A.H.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr. Biol. *13*, 1192–1200.

Li, H., and Stillman, B. (2012). The origin recognition complex: a biochemical and structural view. Subcell. Biochem. *62*, 37–58.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, P.-C., Chretien, L., Côté, J., Kelly, T.J., and Forsburg, S.L. (2011). S. pombe replication protein Cdc18 (Cdc6) interacts with Swi6 (HP1) heterochromatin protein: region specific effects and replication timing in the centromere. Cell Cycle *10*, 323–336.

Li, Y., Bachant, J., Alcasabas, A.A., Wang, Y., Qin, J., and Elledge, S.J. (2002). The mitotic spindle is required for loading of the DASH complex onto the kinetochore. Genes Dev. *16*, 183–197.

Lipford, J.R., and Bell, S.P. (2001). Nucleosomes positioned by ORC facilitate the initiation of DNA replication. Mol. Cell *7*, 21–30.

Liu, J., McConnell, K., Dixon, M., and Calvi, B.R. (2012). Analysis of model replication origins in Drosophila reveals new aspects of the chromatin landscape and its relationship to origin activity and the prereplicative complex. Mol. Biol. Cell *23*, 200–212.

Liu, X., McLeod, I., Anderson, S., Yates, J.R., and He, X. (2005). Molecular analysis of kinetochore architecture in fission yeast. EMBO J. *24*, 2919–2930.

Lo, A.W., Craig, J.M., Saffery, R., Kalitsis, P., Irvine, D. V, Earle, E., Magliano, D.J., and Choo, K.H. (2001). A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. EMBO J. *20*, 2087–2096.

Locovei, A.M., Spiga, M.-G., Tanaka, K., Murakami, Y., and D'Urso, G. (2006). The CENP-B homolog, Abp1, interacts with the initiation protein Cdc23 (MCM10) and is required for efficient DNA replication in fission yeast. Cell Div. *1*, 27.

Lõoke, M., Reimand, J., Sedman, T., Sedman, J., Järvinen, L., Värv, S., Peil, K., Kristjuhan, K., Vilo, J., and Kristjuhan, A. (2010). Relicensing of transcriptionally inactivated replication origins in budding yeast. J. Biol. Chem. *285*, 40004–40011.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.

MacAlpine, D.M., Rodríguez, H.K., and Bell, S.P. (2004). Coordination of replication and transcription along a Drosophila chromosome. Genes Dev. *18*, 3094–3105.

MacAlpine, H.K., Gordân, R., Powell, S.K., Hartemink, A.J., and MacAlpine, D.M. (2010). Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res. *20*, 201–211.

Malik, H. (2002). Conflict begets complexity: the evolution of centromeres. Curr. Opin. Genet. Dev. *12*, 711–718.

Malik, H.S., and Henikoff, S. (2001). Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics *157*, 1293–1298.

Malik, H.S., Vermaak, D., and Henikoff, S. (2002). Recurrent evolution of DNA-binding motifs in the Drosophila centromeric histone. Proc. Natl. Acad. Sci. U. S. A. *99*, 1449–1454.

Malvezzi, F., Litos, G., Schleiffer, A., Heuck, A., Mechtler, K., Clausen, T., and Westermann, S. (2013). A structural basis for kinetochore recruitment of the Ndc80 complex via two distinct centromere receptors. EMBO J. *32*, 409–423.

Marshall, O.J., Chueh, A.C., Wong, L.H., and Choo, K.H.A. (2008). Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am. J. Hum. Genet. *82*, 261–282.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. *9*, 843–854.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., and Okazaki, T. (1989a). A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J. Cell Biol. *109*, 1963–1973.

Masumoto, H., Sugimoto, K., and Okazaki, T. (1989b). Alphoid satellite DNA is tightly associated with centromere antigens in human chromosomes throughout the cell cycle. Exp. Cell Res. *181*, 181–196.

Matsunaga, F., Forterre, P., Ishino, Y., and Myllykallio, H. (2001). In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. Proc. Natl. Acad. Sci. U. S. A. *98*, 11152–11157.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. *18*, 1073–1083.

Mayan, M.D. (2013). RNAP-II Molecules Participate in the Anchoring of the ORC to rDNA Replication Origins. PLoS One *8*, e53405.

McCarroll, R.M., and Fangman, W.L. (1988). Time of replication of yeast centromeres and telomeres. Cell *54*, 505–513.

McCleland, M.L., Kallio, M.J., Barrett-Wilt, G.A., Kestner, C.A., Shabanowitz, J., Hunt, D.F., Gorbsky, G.J., and Stukenberg, P.T. (2004). The vertebrate Ndc80 complex contains Spc24 and Spc25 homologs, which are required to establish and maintain kinetochore-microtubule attachment. Curr. Biol. *14*, 131–137.

McEwen, B.F., Heagle, A.B., Cassels, G.O., Buttle, K.F., and Rieder, C.L. (1997). Kinetochore fiber maturation in PtK1 cells and its implications for the mechanisms of chromosome congression and anaphase onset. J. Cell Biol. *137*, 1567–1580.

McEwen, B.F., Chan, G.K., Zubrowski, B., Savoian, M.S., Sauer, M.T., and Yen, T.J. (2001). CENP-E is essential for reliable bioriented spindle attachment, but chromosome alignment can be achieved via redundant mechanisms in mammalian cells. Mol. Biol. Cell *12*, 2776–2789.

McNairn, A.J., Okuno, Y., Misteli, T., and Gilbert, D.M. (2005). Chinese hamster ORC subunits dynamically associate with chromatin throughout the cell-cycle. Exp. Cell Res. *308*, 345–356.

Melters, D.P., Paliulis, L. V, Korf, I.F., and Chan, S.W.L. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. Chromosome Res.

Meluh, P.B., and Koshland, D. (1995). Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C. Mol. Biol. Cell *6*, 793–807.

Meluh, P.B., and Koshland, D. (1997). Budding yeast centromere composition and assembly as revealed by in vivo cross-linking. Genes Dev. *11*, 3401–3412.

Mendiburo, M.J., Padeken, J., Fülöp, S., Schepers, A., and Heun, P. (2011). Drosophila CENH3 is sufficient for centromere formation. Science *334*, 686–690.

Meyer, C.A., and Liu, X.S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat. Rev. Genet. *15*, 709–721.

Miell, M.D.D., Fuller, C.J., Guse, A., Barysz, H.M., Downes, A., Owen-Hughes, T., Rappsilber, J., Straight, A.F., and Allshire, R.C. (2013). CENP-A confers a reduction in height on octameric nucleosomes. Nat. Struct. Mol. Biol. *20*, 763–765.

Milks, K.J., Moree, B., and Straight, A.F. (2009). Dissection of CENP-C-directed centromere and kinetochore assembly. Mol. Biol. Cell *20*, 4246–4255.

Mishra, P.K., Baum, M., and Carbon, J. (2007). Centromere size and position in Candida albicans are evolutionarily conserved independent of DNA sequence heterogeneity. Mol. Genet. Genomics *278*, 455–465.

Mitchell, A.R., Gosden, J.R., and Miller, D.A. (1985). A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. Chromosoma *92*, 369–377.

Mitchell, A.R., Nicol, L., Malloy, P., and Kipling, D. (1993). Novel structural organisation of a Mus musculus DBA/2 chromosome shows a fixed position for the centromere. J. Cell Sci. *106 (Pt 1*, 79–85.

Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M.M., and Wu, C. (2007). Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. Cell *129*, 1153–1164.

Mojardín, L., Vázquez, E., and Antequera, F. (2013). Specification of DNA Replication Origins and Genomic Base Composition in Fission Yeasts. J. Mol. Biol. *425*, 4706–4713.

Monen, J., Maddox, P.S., Hyndman, F., Oegema, K., and Desai, A. (2005). Differential role of CENP-A in the segregation of holocentric C. elegans chromosomes during meiosis and mitosis. Nat. Cell Biol. *7*, 1248–1255.

Moree, B., Meyer, C.B., Fuller, C.J., and Straight, A.F. (2011). CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. J. Cell Biol. *194*, 855–871.

Moretti, P., and Shore, D. (2001). Multiple interactions in Sir protein recruitment by Rap1p at silencers and telomeres in yeast. Mol. Cell. Biol. *21*, 8082–8094.

Morey, L., Barnes, K., Chen, Y., Fitzgerald-Hayes, M., and Baker, R.E. (2004). The histone fold domain of Cse4 is sufficient for CEN targeting and propagation of active centromeres in budding yeast. Eukaryot. Cell *3*, 1533–1543.

Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789–802.

Muchardt, C., Guilleme, M., Seeler, J.-S., Trouche, D., Dejean, A., and Yaniv, M. (2002). Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1 . EMBO Rep. *3*, 975–981.

Müller, S., Montes de Oca, R., Lacoste, N., Dingli, F., Loew, D., and Almouzni, G. (2014). Phosphorylation and DNA binding of HJURP determine its centromeric recruitment and function in CenH3(CENP-A) loading. Cell Rep. *8*, 190–203.

Murakami, S., Matsumoto, T., Niwa, O., and Yanagida, M. (1991). Structure of the fission yeast centromere cen3: direct analysis of the reiterated inverted region. Chromosoma *101*, 214–221.

Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S., and Hannon, G.J. (2005). Characterization of Dicer-deficient murine embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. *102*, 12135–12140.

Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M., and Okazaki, T. (1992). Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. J. Cell Biol. *116*, 585–596.

Murphy, M.R., Fowlkes, D.M., and Fitzgerald-Hayes, M. (1991). Analysis of centromere function in Saccharomyces cerevisiae using synthetic centromere mutants. Chromosoma *101*, 189–197.

Mythreye, K., and Bloom, K.S. (2003). Differential kinetochore protein requirements for establishment versus propagation of centromere activity in Saccharomyces cerevisiae. J. Cell Biol. *160*, 833–843.

Nakagawa, H., Lee, J.-K., Hurwitz, J., Allshire, R.C., Nakayama, J.-I., Grewal, S.I.S., Tanaka, K., and Murakami, Y. (2002). Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. Genes Dev. *16*, 1766–1778.

Nakano, M., Okamoto, Y., Ohzeki, J., and Masumoto, H. (2003). Epigenetic assembly of centromeric chromatin at ectopic alpha-satellite sites on human chromosomes. J. Cell Sci. *116*, 4021–4034.

Nakaseko, Y., Kinoshita, N., and Yanagida, M. (1987). A novel sequence common to the centromere regions of Schizosaccharomyces pombe chromosomes. Nucleic Acids Res. *15*, 4705–4715.

Natsume, T., Müller, C.A., Katou, Y., Retkute, R., Gierliński, M., Araki, H., Blow, J.J., Shirahige, K., Nieduszynski, C.A., and Tanaka, T.U. (2013). Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. Mol. Cell *50*, 661–674.

Neumann, P., Navrátilová, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerová, V., Chocholová, E., Novák, P., Wanner, G., and Macas, J. (2012). Stretching the rules: monocentric chromosomes with multiple centromere domains. PLoS Genet. *8*, e1002777.

Neumann, P., Pavlikova, Z., Kobli kova, A., Fukova, I., Jedli kova, V., Novak, P., and Macas, J. (2015). Centromeres off the hook: massive changes in centromere size and structure following duplication of CenH3 gene in Fabeae species. Mol. Biol. Evol.

Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E. V. (1999). AAA+: A Class of Chaperone-Like ATPases Associated with the Assembly, Operation, and Disassembly of Protein Complexes. Genome Res. *9*, 27–43.

Nishino, T., Takeuchi, K., Gascoigne, K.E., Suzuki, A., Hori, T., Oyama, T., Morikawa, K., Cheeseman, I.M., and Fukagawa, T. (2012). CENP-T-W-S-X Forms a Unique Centromeric Chromatin Structure with a Histone-like Fold. Cell *148*, 487–501.

Nonaka, N., Kitajima, T., Yokobayashi, S., Xiao, G., Yamamoto, M., Grewal, S.I.S., and Watanabe, Y. (2002). Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. Nat. Cell Biol. *4*, 89–93.

Obado, S.O., Bot, C., Nilsson, D., Andersson, B., and Kelly, J.M. (2007). Repetitive DNA is associated with centromeric domains in Trypanosoma brucei but not Trypanosoma cruzi. Genome Biol. *8*, R37.

Obuse, C., Yang, H., Nozaki, N., Goto, S., Okazaki, T., and Yoda, K. (2004). Proteomics analysis of the centromere complex from HeLa interphase cells: UV-

damaged DNA binding protein 1 (DDB-1) is a component of the CEN-complex, while BMI-1 is transiently co-localized with the centromeric region in interphase. Genes Cells *9*, 105–120.

Ogiyama, Y., Ohno, Y., Kubota, Y., and Ishii, K. (2013). Epigenetically induced paucity of histone H2A.Z stabilizes fission-yeast ectopic centromeres. Nat. Struct. Mol. Biol.

Ohkuni, K., and Kitagawa, K. (2011). Endogenous transcription at the centromere facilitates centromere activity in budding yeast. Curr. Biol. *21*, 1695–1703.

Ohzeki, J., Nakano, M., Okada, T., and Masumoto, H. (2002). CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J. Cell Biol. *159*, 765–775.

Okada, M., Cheeseman, I.M., Hori, T., Okawa, K., McLeod, I.X., Yates, J.R., Desai, A., and Fukagawa, T. (2006). The CENP-H-I complex is required for the efficient incorporation of newly synthesized CENP-A into centromeres. Nat. Cell Biol. *8*, 446–457.

Okada, M., Okawa, K., Isobe, T., and Fukagawa, T. (2009). CENP-H-containing complex facilitates centromere deposition of CENP-A in cooperation with FACT and CHD1. Mol. Biol. Cell *20*, 3986–3995.

Okada, T., Ohzeki, J., Nakano, M., Yoda, K., Brinkley, W.R., Larionov, V., and Masumoto, H. (2007). CENP-B controls centromere formation depending on the chromatin context. Cell *131*, 1287–1300.

Olson, W.K., Gorin, A.A., Lu, X.-J.J., Hock, L.M., and Zhurkin, V.B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc. Natl. Acad. Sci. U. S. A. *95*, 11163–11168.

Padeganeh, A., Ryan, J., Boisvert, J., Ladouceur, A.-M., Dorn, J.F., and Maddox, P.S. (2013). Octameric CENP-A nucleosomes are present at human centromeres throughout the cell cycle. Curr. Biol. *23*, 764–769.

Padmanabhan, S., Thakur, J., Siddharthan, R., and Sanyal, K. (2008). Rapid evolution of Cse4p-rich centromeric DNA sequences in closely related pathogenic yeasts, Candida albicans and Candida dubliniensis. Proc. Natl. Acad. Sci. U. S. A. *105*, 19797–19802.

Pak, D.T., Pflumm, M., Chesnokov, I., Huang, D.W., Kellum, R., Marr, J., Romanowski, P., and Botchan, M.R. (1997). Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. Cell *91*, 311–323.

Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Chikka, M.R., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C.R. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669–672.

Palmer, D.K., O'Day, K., Wener, M.H., Andrews, B.S., and Margolis, R.L. (1987). A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J. Cell Biol. *104*, 805–815.

Palmer, D.K., O'Day, K., and Margolis, R.L. (1990). The centromere specific histone CENP-A is selectively retained in discrete foci in mammalian sperm nuclei. Chromosoma *100*, 32–36.

Palmer, D.K., O'Day, K., Trong, H.L., Charbonneau, H., and Margolis, R.L. (1991). Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. Proc. Natl. Acad. Sci. U. S. A. *88*, 3734–3738.

Partridge, J.F., Borgstrom, B., Allshire, R.C., Borgstrøm, B., and Allshire, R.C. (2000). Distinct protein interaction domains and protein spreading in a complex centromere. Genes Dev. *14*, 783–791.

Petrovic, A., Pasqualato, S., Dube, P., Krenn, V., Santaguida, S., Cittaro, D., Monzani, S., Massimiliano, L., Keller, J., Tarricone, A., et al. (2010). The MIS12 complex is a protein interaction hub for outer kinetochore assembly. J. Cell Biol. *190*, 835–852.

Pidoux, A.L., Richardson, W., and Allshire, R.C. (2003). Sim4: a novel fission yeast kinetochore protein required for centromeric silencing and chromosome segregation. J. Cell Biol. *161*, 295–307.

Pidoux, A.L., Choi, E.S., Abbott, J.K.R., Liu, X., Kagansky, A., Castillo, A.G., Hamilton, G.L., Richardson, W., Rappsilber, J., He, X., et al. (2009). Fission yeast Scm3: A CENP-A receptor required for integrity of subkinetochore chromatin. Mol. Cell *33*, 299–311.

Pinsky, B.A., Tatsutani, S.Y., Collins, K.A., and Biggins, S. (2003). An Mtw1 Complex Promotes Kinetochore Biorientation that Is Monitored by the Ipl1/Aurora Protein Kinase. Dev. Cell *5*, 735–745.

Piras, F.M., Nergadze, S.G., Magnani, E., Bertoni, L., Attolini, C., Khoriauli, L., Raimondi, E., and Giulotto, E. (2010). Uncoupling of satellite DNA and centromeric function in the genus Equus. PLoS Genet. *6*, e1000845.

Pohl, T.J., Brewer, B.J., and Raghuraman, M.K. (2012). Functional Centromeres Determine the Activation Time of Pericentric Origins of DNA Replication in Saccharomyces cerevisiae. PLoS Genet. *8*, e1002677.

Powell, S.K., MacAlpine, H.K., Prinz, J.A., Li, Y., Belsky, J.A., and MacAlpine, D.M. (2015). Dynamic loading and redistribution of the Mcm2-7 helicase complex through the cell cycle. EMBO J. *34*, 531–543.

Powers, A.F., Franck, A.D., Gestaut, D.R., Cooper, J., Gracyzk, B., Wei, R.R., Wordeman, L., Davis, T.N., and Asbury, C.L. (2009). The Ndc80 kinetochore complex forms load-bearing attachments to dynamic microtubule tips via biased diffusion. Cell *136*, 865–875.

Prasanth, S.G., Prasanth, K. V, Siddiqui, K., Spector, D.L., and Stillman, B. (2004). Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance. EMBO J. *23*, 2651–2663.

Prasanth, S.G., Shen, Z., Prasanth, K. V, and Stillman, B. (2010). Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization. Proc. Natl. Acad. Sci. U. S. A. *107*, 15093–15098.

Prendergast, L., van Vuuren, C., Kaczmarczyk, A., Doering, V., Hellwig, D., Quinn, N., Hoischen, C., Diekmann, S., and Sullivan, K.F. (2011). Premitotic assembly of human CENPs -T and -W switches centromeric chromatin to a mitotic state. PLoS Biol. *9*, e1001082.

Prioleau, M.-N., Gendron, M.-C., and Hyrien, O. (2003). Replication of the chicken beta-globin locus: early-firing origins at the 5' HS4 insulator and the rho- and betaA-globin genes show opposite epigenetic modifications. Mol. Cell. Biol. *23*, 3536–3549.

Prokhorova, T.A., and Blow, J.J. (2000). Sequential MCM/P1 Subcomplex Assembly Is Required to Form a Heterohexamer with Replication Licensing Activity. J. Biol. Chem. *275*, 2491–2498.

Prunell, A. (1982). Nucleosome reconstitution on plasmid-inserted poly(dA) . poly(dT). EMBO J. *1*, 173–179.

Przewloka, M.R., Venkei, Z., Bolanos-Garcia, V.M., Debski, J., Dadlez, M., and Glover, D.M. (2011). CENP-C is a structural platform for kinetochore assembly. Curr. Biol. *21*, 399–405.

Purgato, S., Belloni, E., Piras, F.M., Zoli, M., Badiale, C., Cerutti, F., Mazzagatti, A., Perini, G., Della Valle, G., Nergadze, S.G., et al. (2014). Centromere sliding on a mammalian chromosome. Chromosoma.

Rago, F., Gascoigne, K.E., and Cheeseman, I.M. (2015). Distinct Organization and Regulation of the Outer Kinetochore KMN Network Downstream of CENP-C and CENP-T. Curr. Biol. *25*, 671–677.

Ranjitkar, P., Press, M.O., Yi, X., Baker, R., MacCoss, M.J., and Biggins, S. (2010). An E3 ubiquitin ligase prevents ectopic localization of the centromeric histone H3 variant via the centromere targeting domain. Mol. Cell *40*, 455–464.

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., and Segal, E. (2012). Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat. Genet. *44*, 743–750.

Rehman, M.A., Fourel, G., Mathews, A., Ramdin, D., Espinosa, M., Gilson, E., and Yankulov, K. (2006). Differential requirement of DNA replication factors for subtelomeric ARS consensus sequence protosilencers in Saccharomyces cerevisiae. Genetics *174*, 1801–1810.

Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

Remus, D., Beall, E.L., and Botchan, M.R. (2004). DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. EMBO J. *23*, 897–907.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F.X. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell *139*, 719–730.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408–1419.

Rhind, N., Chen, Z., Yassour, M., Thompson, D.A., Haas, B.J., Habib, N., Wapinski, I., Roy, S., Lin, M.F., Heiman, D.I., et al. (2011). Comparative functional genomics of the fission yeasts. Science *332*, 930–936.

Ribeiro, S.A., Vagnarelli, P., Dong, Y., Hori, T., McEwen, B.F., Fukagawa, T., Flors, C., and Earnshaw, W.C. (2010). A super-resolution map of the vertebrate kinetochore. Proc. Natl. Acad. Sci. U. S. A. *107*, 10484–10489.

Ritzi, M. (1998). Human Minichromosome Maintenance Proteins and Human Origin Recognition Complex 2 Protein on Chromatin. J. Biol. Chem. *273*, 24543–24549.

Robbins, E., and Gonatas, N.K. (1964). THE ULTRASTRUCTURE OF A MAMMALIAN CELL DURING THE MITOTIC CYCLE. J. Cell Biol. *21*, 429–463.

Rocchi, M., Archidiacono, N., Schempp, W., Capozzi, O., and Stanyon, R. (2012). Centromere repositioning in mammals. Heredity (Edinb). *108*, 59–67.

Roos, U.P. (1973). Light and electron microscopy of rat kangaroo cells in mitosis. II. Kinetochore structure and function. Chromosoma *41*, 195–220.

Rusché, L.N., Kirchmaier, A.L., and Rine, J. (2002). Ordered nucleation and spreading of silenced chromatin in Saccharomyces cerevisiae. Mol. Biol. Cell *13*, 2207–2222.

Saffery, R., Irvine, D. V, Griffiths, B., Kalitsis, P., Wordeman, L., and Choo, K.H. (2000). Human centromeres and neocentromeres show identical distribution patterns of >20 functionally important kinetochore-associated proteins. Hum. Mol. Genet. *9*, 175–185.

Saffery, R., Sumer, H., Hassan, S., Wong, L.H., Craig, J.M., Todokoro, K., Anderson, M., Stafford, A., and Choo, K.H.A. (2003). Transcription within a functional human centromere. Mol. Cell *12*, 509–516.

Saitoh, H., Tomkiel, J., Cooke, C.A., Ratrie, H., Maurer, M., Rothfield, N.F., and Earnshaw, W.C. (1992). CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. Cell *70*, 115–125.

Sanchez-Perez, I., Renwick, S.J., Crawley, K., Karig, I., Buck, V., Meadows, J.C., Franco-Sanchez, A., Fleig, U., Toda, T., and Millar, J.B.A. (2005). The DASH complex and Klp5/Klp6 kinesin coordinate bipolar chromosome attachment in fission yeast. EMBO J. *24*, 2931–2943.

Sanchez-Pulido, L., Pidoux, A.L., Ponting, C.P., and Allshire, R.C. (2009). Common ancestry of the CENP-A chaperones Scm3 and HJURP. Cell *137*, 1173–1174.

Sanyal, K., and Carbon, J. (2002). The CENP-A homolog CaCse4p in the pathogenic yeast Candida albicans is a centromere protein essential for chromosome transmission. Proc. Natl. Acad. Sci. U. S. A. *99*, 12969–12974.

Sanyal, K., Baum, M., and Carbon, J. (2004). Centromeric DNA sequences in the pathogenic yeast Candida albicans are all different and unique. Proc. Natl. Acad. Sci. U. S. A. *101*, 11374–11379.

Sart, D. du, Cancilla, M.R., Earle, E., Mao, J.I., Saffery, R., Tainton, K.M., Kalitsis, P., Martyn, J., Barry, A.E., Choo, K.H.A., et al. (1997). A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. Nat. Genet. *16*, 144–153.

Sato, H., and Saitoh, S. (2013). Switching the centromeres on and off: epigenetic chromatin alterations provide plasticity in centromere activity stabilizing aberrant dicentric chromosomes. Biochem. Soc. Trans. *41*, 1648–1653.

Sato, H., Masuda, F., Takayama, Y., Takahashi, K., and Saitoh, S. (2012). Epigenetic inactivation and subsequent heterochromatinization of a centromere stabilize dicentric chromosomes. Curr. Biol. *22*, 658–667.

Schieferstein, U., and Thoma, F. (1996). Modulation of cyclobutane pyrimidine dimer formation in a positioned nucleosome containing poly(dA.dT) tracts. Biochemistry *35*, 7705–7714.

Schleiffer, A., Maier, M., Litos, G., Lampert, F., Hornung, P., Mechtler, K., and Westermann, S. (2012). CENP-T proteins are conserved centromere receptors of the Ndc80 complex. Nat. Cell Biol. *14*, 1–12.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887–898.

Schueler, M.G., Higgins, A.W., Rudd, M.K., Gustashaw, K., and Willard, H.F. (2001). Genomic and genetic definition of a functional human centromere. Science *294*, 109–115.

Schueler, M.G., Swanson, W., Thomas, P.J., and Green, E.D. (2010). Adaptive evolution of foundation kinetochore proteins in primates. Mol. Biol. Evol. *27*, 1585–1597.

Schultz, S.S., Desbordes, S.C., Du, Z., Kosiyatrakul, S., Lipchina, I., Studer, L., and Schildkraut, C.L. (2010). Single-molecule analysis reveals changes in the DNA

replication program for the POU5F1 locus upon human embryonic stem cell differentiation. Mol. Cell. Biol. *30*, 4521–4534.

Schwacha, A., and Bell, S.P. (2001). Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication. Mol. Cell *8*, 1093–1104.

Scott, K.C. (2013). Transcription and ncRNAs: at the cent(rome)re of kinetochore assembly and maintenance. Chromosome Res. *21*, 643–651.

Scott, K.C., and Sullivan, B.A. (2014). Neocentromeres: a place for everything and everything in its place. Trends Genet. *30*, 66–74.

Scott, K.C., Merrett, S.L., and Willard, H.F. (2006). A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. Curr. Biol. *16*, 119–129.

Scott, K.C., White, C. V, and Willard, H.F. (2007). An RNA polymerase III-dependent heterochromatin barrier at fission yeast centromere 1. PLoS One *2*, e1099.

Screpanti, E., De Antoni, A., Alushin, G.M., Petrovic, A., Melis, T., Nogales, E., and Musacchio, A. (2011). Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. Curr. Biol. *21*, 391–398.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I.K., Wang, J.-P.Z., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772–778.

Segurado, M., de Luis, A., and Antequera, F. (2003). Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe. EMBO Rep. *4*, 1048–1053.

Sequeira-Mendes, J., Díaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gómez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet. *5*, e1000446.

Shang, W.-H., Hori, T., Toyoda, A., Kato, J., Popendorf, K., Sakakibara, Y., Fujiyama, A., and Fukagawa, T. (2010). Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. Genome Res. *20*, 1219–1228.

Shang, W.-H., Hori, T., Martins, N.M.C., Toyoda, A., Misu, S., Monma, N., Hiratani, I., Maeshima, K., Ikeo, K., Fujiyama, A., et al. (2013). Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. Dev. Cell *24*, 635–648.

Shankaranarayana, G.D., Motamedi, M.R., Moazed, D., and Grewal, S.I.S.S. (2003). Sir2 Regulates Histone H3 Lysine 9 Methylation and Heterochromatin Assembly in Fission Yeast. Curr. Biol. *13*, 1240–1246.

Shanker, S., Job, G., George, O.L., Creamer, K.M., Shaban, A., and Partridge, J.F. (2010). Continuous requirement for the Clr4 complex but not RNAi for centromeric heterochromatin assembly in fission yeast harboring a disrupted RITS complex. PLoS Genet. *6*, e1001174.

Shareef, M.M., King, C., Damaj, M., Badagu, R., Huang, D.W., and Kellum, R. (2001). Drosophila heterochromatin protein 1 (HP1)/origin recognition complex (ORC) protein is associated with HP1 and ORC and functions in heterochromatin-induced silencing. Mol. Biol. Cell *12*, 1671–1685.

Shareef, M.M., Badugu, R., and Kellum, R. (2003). HP1/ORC complex and heterochromatin assembly. Genetica *117*, 127–134.

Shelby, R.D., Vafa, O., and Sullivan, K.F. (1997). Assembly of CENP-A into Centromeric Chromatin Requires a Cooperative Array of Nucleosomal DNA Contact Sites. J. Cell Biol. *136*, 501–513.

Shelby, R.D., Monier, K., and Sullivan, K.F. (2000). Chromatin assembly at kinetochores is uncoupled from DNA replication. J. Cell Biol. *151*, 1113–1118.

Shivaraju, M., Unruh, J.R.R., Slaughter, B.D.D., Mattingly, M., Berman, J., and Gerton, J.L.L. (2012). Cell-Cycle-Coupled Structural Oscillation of Centromeric Nucleosomes in Yeast. Cell *150*, 304–316.

Shuaib, M., Ouararhni, K., Dimitrov, S., and Hamiche, A. (2010). HJURP binds CENP-A via a highly conserved N-terminal domain and mediates its deposition at centromeres. Proc. Natl. Acad. Sci. U. S. A. *107*, 1349–1354.

Silva, M.C.C., Bodor, D.L., Stellfox, M.E., Martins, N.M.C., Hochegger, H., Foltz, D.R., and Jansen, L.E.T. (2012). Cdk activity couples epigenetic centromere inheritance to cell cycle progression. Dev. Cell *22*, 52–63.

Simic, R., Lindstrom, D.L., Tran, H.G., Roinick, K.L., Costa, P.J., Johnson, A.D., Hartzog, G.A., and Arndt, K.M. (2003). Chromatin remodeling protein Chd1 interacts with transcription elongation factors and localizes to transcribed genes. EMBO J. *22*, 1846–1856.

Singh, T.R., Saro, D., Ali, A.M., Zheng, X.-F., Du, C., Killen, M.W., Sachpatzidis, A., Wahengbam, K., Pierce, A.J., Xiong, Y., et al. (2010). MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM. Mol. Cell *37*, 879–886.

Siow, C.C., Nieduszynska, S.R., Müller, C.A., and Nieduszynski, C.A. (2012). OriDB, the DNA replication origin database updated and extended. Nucleic Acids Res. *40*, D682–D686.

Smith, J., Caddle, M., Bulboaca, G., Wohlgemuth, J., Baum, M., Clarke, L., and Calos, M. (1995). Replication of centromere II of Schizosaccharomyces pombe. Mol. Cell. Biol. *15*, 5165–5172.

Song, J.S., Liu, X., Liu, X.S., and He, X. (2008). A high-resolution map of nucleosome positioning on a fission yeast centromere. Genome Res. *18*, 1064–1072.

Steiner, F.A., and Henikoff, S. (2014). Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. Elife *3*, e02025.

Stimpson, K.M., Song, I.Y., Jauch, A., Holtgreve-Grez, H., Hayden, K.E., Bridger, J.M., and Sullivan, B.A. (2010). Telomere disruption results in non-random formation of de novo dicentric chromosomes involving acrocentric human chromosomes. PLoS Genet. *6*.

Stoler, S., Keith, K.C., Curnick, K.E., and Fitzgerald-Hayes, M. (1995). A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev. *9*, 573–586.

Stoler, S., Rogers, K., Weitze, S., Morey, L., Fitzgerald-Hayes, M., and Baker, R.E. (2007). Scm3, an essential Saccharomyces cerevisiae centromere protein required for G2/M progression and Cse4 localization. Proc. Natl. Acad. Sci. U. S. A. *104*, 10571–10576.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. *20*, 267–273.

Subramanian, L., Toda, N.R.T., Rappsilber, J., and Allshire, R.C. (2014). Eic1 links Mis18 with the CCAN/Mis6/Ctf19 complex to promote CENP-A assembly. Open Biol. *4*, 140043.

Sugata, N., Munekata, E., and Todokoro, K. (1999). Characterization of a novel kinetochore protein, CENP-H. J. Biol. Chem. *274*, 27343–27346.

Sugiyama, T., Cam, H., Verdel, A., Moazed, D., and Grewal, S.I.S. (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. Proc. Natl. Acad. Sci. U. S. A. *102*, 152–157.

Sullivan, B., and Karpen, G. (2001). Centromere identity in Drosophila is not determined in vivo by replication timing. J. Cell Biol. *154*, 683–690.

Sullivan, B.A., and Schwartz, S. (1995). Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. Hum. Mol. Genet. *4*, 2189–2197.

Sullivan, B.A., and Willard, H.F. (1998). Stable dicentric X chromosomes with two functional centromeres. Nat. Genet. *20*, 227–228.

Sullivan, K.F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J. Cell Biol. *127*, 581–592.

Sun, X., Le, H.D., Wahlstrom, J.M., and Karpen, G.H. (2003). Sequence analysis of a functional Drosophila centromere. Genome Res. *13*, 182–194.

Surosky, R.T., and Tye, B.K. (1985). Resolution of dicentric chromosomes by Ty-mediated recombination in yeast. Genetics *110*, 397–419.

Sutton, A., Heller, R.C., Landry, J., Choy, J.S., Sirko, A., and Sternglanz, R. (2001). A novel form of transcriptional silencing by Sum1-1 requires Hst1 and the origin recognition complex. Mol. Cell. Biol. *21*, 3514–3522.

Suzuki, A., Hori, T., Nishino, T., Usukura, J., Miyagi, A., Morikawa, K., and Fukagawa, T. (2011). Spindle microtubules generate tension-dependent changes in the distribution of inner kinetochore proteins. J. Cell Biol. *193*, 125–140.

Tachiwana, H., Osakabe, A., Shiga, T., Miya, Y., Kimura, H., Kagawa, W., and Kurumizaka, H. (2011a). Structures of human nucleosomes containing major histone H3 variants. Acta Crystallogr. D. Biol. Crystallogr. *67*, 578–583.

Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.-Y., et al. (2011b). Crystal structure of the human centromeric nucleosome containing CENP-A. Nature *476*, 232–235.

Tachiwana, H., Kagawa, W., and Kurumizaka, H. (2012). Comparison between the CENP-A and histone H3 structures in nucleosomes. Nucleus *3*.

Tachiwana, H., Müller, S., Blümer, J., Klare, K., Musacchio, A., and Almouzni, G. (2015). HJURP Involvement in De Novo CenH3(CENP-A) and CENP-C Recruitment. Cell Rep. *11*, 22–32.

Takahashi, K., Murakami, S., Chikashige, Y., Niwa, O., and Yanagida, M. (1991). A large number of tRNA genes are symmetrically located in fission yeast centromeres. J. Mol. Biol. *218*, 13–17.

Takahashi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O., and Yanagida, M. (1992). A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. Mol. Biol. Cell *3*, 819–835.

Takahashi, K., Chen, E.S., and Yanagida, M. (2000). Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. Science (80-.). *288*, 2215–2219.

Takahashi, K., Takayama, Y., Masuda, F., Kobayashi, Y., and Saitoh, S. (2005). Two distinct pathways responsible for the loading of CENP-A to centromeres in the fission yeast cell cycle. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *360*, 595–606; discussion 606–607.

Takahashi, T., Ohara, E., Nishitani, H., and Masukata, H. (2003). Multiple ORC-binding sites are required for efficient MCM loading and origin firing in fission yeast. EMBO J. *22*, 964–974.

Takara, T.J., and Bell, S.P. (2011). Multiple Cdt1 molecules act at each origin to load replication-competent Mcm2-7 helicases. EMBO J. *30*, 4885–4896.

Takayama, Y., Sato, H., Saitoh, S., Ogiyama, Y., Masuda, F., and Takahashi, K. (2008). Biphasic incorporation of centromeric histone CENP-A in fission yeast. Mol. Biol. Cell *19*, 682–690.

Takeuchi, K., and Fukagawa, T. (2012). Molecular architecture of vertebrate kinetochores. Exp. Cell Res. *318*, 1367–1374.

Talbert, P.B., Masuelli, R., Tyagi, A.P., Comai, L., and Henikoff, S. (2002). Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. Plant Cell *14*, 1053–1066.

Tanaka, S., and Araki, H. (2010). Regulation of the initiation step of DNA replication by cyclin-dependent kinases. Chromosoma *119*, 565–574.

Tanaka, Y., Nureki, O., Kurumizaka, H., Fukai, S., Kawaguchi, S., Ikuta, M., Iwahara, J., Okazaki, T., and Yokoyama, S. (2001). Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. EMBO J. *20*, 6612–6618.

Tanny, J.C., Dowd, G.J., Huang, J., Hilz, H., and Moazed, D. (1999). An Enzymatic Activity in the Yeast Sir2 Protein that Is Essential for Gene Silencing. Cell *99*, 735–745.

Tao, Y., Jin, C., Li, X., Qi, S., Chu, L., Niu, L., Yao, X., and Teng, M. (2012). The structure of the FANCM-MHF complex reveals physical features for functional assembly. Nat. Commun. *3*, 782.

Team, R.C. (2013). R: A language and environment for statistical computing. R Found. Stat. Comput. Vienna, Austria *ISBN 3-900*.

Thakur, J., and Sanyal, K. (2013). Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function at native physical chromosomal loci in Candida albicans. Genome Res. *23*, 638–652.

Thåström, A., Lowary, P.T., Widlund, H.R., Cao, H., Kubista, M., and Widom, J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. J. Mol. Biol. *288*, 213–229.

Thomae, A.W., Pich, D., Brocher, J., Spindler, M.-P., Berens, C., Hock, R., Hammerschmidt, W., and Schepers, A. (2008). Interaction between HMGA1a and the origin recognition complex creates site-specific replication origins. Proc. Natl. Acad. Sci. U. S. A. *105*, 1692–1697.

Thomae, A.W., Baltin, J., Pich, D., Deutsch, M.J., Ravasz, M., Zeller, K., Gossen, M., Hammerschmidt, W., and Schepers, A. (2011). Different roles of the human Orc6 protein in the replication initiation process. Cell. Mol. Life Sci. *68*, 3741–3756.

Tien, J.F., Umbreit, N.T., Gestaut, D.R., Franck, A.D., Cooper, J., Wordeman, L., Gonen, T., Asbury, C.L., and Davis, T.N. (2010). Cooperation of the Dam1 and Ndc80 kinetochore complexes enhances microtubule coupling and is regulated by aurora B. J. Cell Biol. *189*, 713–723.

Tomkiel, J., Cooke, C.A., Saitoh, H., Bernat, R.L., and Earnshaw, W.C. (1994). CENP-C is required for maintaining proper kinetochore size and for a timely transition to anaphase. J. Cell Biol. *125*, 531–545.

Tong, K., Keller, T., Hoffman, C.S., and Annunziato, A.T. (2012). Schizosaccharomyces pombe Hat1 (Kat1) is associated with Mis16 and is required for telomeric silencing. Eukaryot. Cell *11*, 1095–1103.

Topp, C.N., Zhong, C.X., and Dawe, R.K. (2004). Centromere-encoded RNAs are integral components of the maize kinetochore. Proc. Natl. Acad. Sci. U. S. A. *101*, 15986–15991.

Triolo, T., and Sternglanz, R. (1996). Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. Nature *381*, 251–253.

Tsakraklides, V., and Bell, S.P. (2010). Dynamics of pre-replicative complex assembly. J. Biol. Chem. *285*, 9437–9443.

Tsankov, A., Yanagisawa, Y., Rhind, N., Regev, A., and Rando, O.J. (2011). Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. Genome Res. *21*, 1851–1862.

Tsankov, A.M., Thompson, D.A., Socha, A., Regev, A., and Rando, O.J. (2010). The role of nucleosome positioning in the evolution of gene regulation. PLoS Biol. *8*, e1000414.

Tsuyama, T., Tada, S., Watanabe, S., Seki, M., and Enomoto, T. (2005). Licensing for DNA replication requires a strict sequential assembly of Cdc6 and Cdt1 onto chromatin in Xenopus egg extracts. Nucleic Acids Res. *33*, 765–775.

Tyler-Smith, C., Gimelli, G., Giglio, S., Floridia, G., Pandya, A., Terzoli, G., Warburton, P.E., Earnshaw, W.C., and Zuffardi, O. (1999). Transmission of a fully functional human neocentromere through three generations. Am. J. Hum. Genet. *64*, 1440–1444.

Vafa, O., and Sullivan, K.F. (1997). Chromatin containing CENP-A and α-satellite DNA is a major component of the inner kinetochore plate. Curr. Biol. *7*, 897–900.

Vaillant, C., Palmeira, L., Chevereau, G., Audit, B., d'Aubenton-Carafa, Y., Thermes, C., and Arneodo, A. (2010). A novel strategy of transcription regulation by intragenic nucleosome ordering. Genome Res. *20*, 59–67.

Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T.J., and Walter, J.C. (2003). Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes Dev. *17*, 1894–1908.

Venkei, Z., Przewloka, M.R., and Glover, D.M. (2011). Drosophila Mis12 complex acts as a single functional unit essential for anaphase chromosome movement and a robust spindle assembly checkpoint. Genetics *187*, 131–140.

Ventura, M., Mudge, J.M., Palumbo, V., Burn, S., Blennow, E., Pierluigi, M., Giorda, R., Zuffardi, O., Archidiacono, N., Jackson, M.S., et al. (2003). Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. Genome Res. *13*, 2059–2068.

Ventura, M., Weigl, S., Carbone, L., Cardone, M.F., Misceo, D., Teti, M., D'Addabbo, P., Wandall, A., Björck, E., de Jong, P.J., et al. (2004). Recurrent Sites for New Centromere Seeding. Genome Res. *14*, 1696–1703.

Ventura, M., Antonacci, F., Cardone, M.F., Stanyon, R., D'Addabbo, P., Cellamare, A., Sprague, L.J., Eichler, E.E., Archidiacono, N., and Rocchi, M. (2007). Evolutionary formation of new centromeres in macaque. Science *316*, 243–246.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I.S., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672–676.

Vermaak, D., Hayden, H.S., and Henikoff, S. (2002). Centromere targeting element within the histone fold domain of Cid. Mol. Cell. Biol. *22*, 7553–7561.

Visnapuu, M.-L., and Greene, E.C. (2009). Single-molecule imaging of DNA curtains reveals intrinsic energy landscapes for nucleosome deposition. Nat. Struct. Mol. Biol. *16*, 1056–1062.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833–1837.

Voullaire, L.E., Slater, H.R., Petrovic, V., and Choo, K.H.A. (1993). A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? Am. J. Hum. Genet. *52*, 1153–1163.

Wade, C.M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., Lear, T.L., Adelson, D.L., Bailey, E., Bellone, R.R., et al. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. Science *326*, 865–867.

Walfridsson, J., Bjerling, P., Thalen, M., Yoo, E.-J., Park, S.D., and Ekwall, K. (2005). The CHD remodeling factor Hrp1 stimulates CENP-A loading to centromeres. Nucleic Acids Res. *33*, 2868–2879.

Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., Gimelli, G., Warburton, D., Tyler-Smith, C., Sullivan, K.F., et al. (1997). Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. Curr. Biol. *7*, 901–904.

Watanabe, Y., and Nurse, P. (1999). Cohesin Rec8 is required for reductional chromosome segregation at meiosis. Nature *400*, 461–464.

Watanabe, Y., Yokobayashi, S., Yamamoto, M., and Nurse, P. (2001). Pre-meiotic S phase is linked to reductional chromosome segregation and recombination. Nature *409*, 359–363.

Waye, J.S., and Willard, H.F. (1989). Human beta satellite DNA: genomic organization and sequence definition of a class of highly repetitive tandem DNA. Proc. Natl. Acad. Sci. *86*, 6250–6254.

Wei, R.R., Sorger, P.K., and Harrison, S.C. (2005). Molecular organization of the Ndc80 complex, an essential kinetochore component. Proc. Natl. Acad. Sci. U. S. A. *102*, 5363–5367.

Weiner, A., Hughes, A., Yassour, M., Rando, O.J., and Friedman, N. (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res. *20*, 90–100.

Weinreich, M., Liang, C., and Stillman, B. (1999). The Cdc6p nucleotide-binding motif is required for loading mcm proteins onto chromatin. Proc. Natl. Acad. Sci. U. S. A. *96*, 441–446.

Wendell, K.L., Wilson, L., and Jordan, M.A. (1993). Mitotic block in HeLa cells by vinblastine: ultrastructural changes in kinetochore-microtubule attachment and in centrosomes. J. Cell Sci. *104 (Pt 2*, 261–274.

Westermann, S., and Schleiffer, A. (2013). Family matters: structural and functional conservation of centromere-associated proteins from yeast to humans. Trends Cell Biol.

Westermann, S., Avila-Sakar, A., Wang, H.-W., Niederstrasser, H., Wong, J., Drubin, D.G., Nogales, E., and Barnes, G. (2005). Formation of a dynamic kinetochore- microtubule interface through assembly of the Dam1 ring complex. Mol. Cell *17*, 277–290.

Wieland, G., Orthaus, S., Ohndorf, S., Diekmann, S., and Hemmerich, P. (2004). Functional complementation of human centromere protein A (CENP-A) by Cse4p from Saccharomyces cerevisiae. Mol. Cell. Biol. *24*, 6620–6630.

Wigge, P.A., and Kilmartin, J. V (2001). The Ndc80p complex from Saccharomyces cerevisiae contains conserved centromere components and has a function in chromosome segregation. J. Cell Biol. *152*, 349–360.

Willard, H.F. (1985). Chromosome-specific organization of human alpha satellite DNA. Am. J. Hum. Genet. *37*, 524–532.

Williams, J.S., Hayashi, T., Yanagida, M., and Russell, P. (2009). Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. Mol. Cell *33*, 287–298.

Winey, M., Mamay, C.L., O'Toole, E.T., Mastronarde, D.N., Giddings, T.H., McDonald, K.L., and McIntosh, J.R. (1995). Three-dimensional ultrastructural analysis of the Saccharomyces cerevisiae mitotic spindle. J. Cell Biol. *129*, 1601–1615.

Wisniewski, J., Hajj, B., Chen, J., Mizuguchi, G., Xiao, H., Wei, D., Dahan, M., and Wu, C. (2014). Imaging the fate of histone Cse4 reveals de novo replacement in S phase and subsequent stable residence at centromeres. Elife *3*, e02203.

Wong, L.H., Brettingham-Moore, K.H., Chan, L., Quach, J.M., Anderson, M.A., Northrop, E.L., Hannan, R., Saffery, R., Shaw, M.L., Williams, E., et al. (2007). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res. *17*, 1146–1160.

Wood, V., Harris, M.A., McDowall, M.D., Rutherford, K., Vaughan, B.W., Staines, D.M., Aslett, M., Lock, A., Bähler, J., Kersey, P.J., et al. (2012). PomBase: a comprehensive online resource for fission yeast. Nucleic Acids Res. *40*, D695–D699.

Woodward, A.M., Göhler, T., Luciani, M.G., Oehlmann, M., Ge, X., Gartner, A., Jackson, D.A., and Blow, J.J. (2006). Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. J. Cell Biol. *173*, 673–683.

De Wulf, P., McAinsh, A.D., and Sorger, P.K. (2003). Hierarchical assembly of the budding yeast kinetochore from multiple subcomplexes. Genes Dev. *17*, 2902–2921.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science *294*, 2357–2360.

Xiao, H., Mizuguchi, G., Wisniewski, J., Huang, Y., Wei, D., and Wu, C. (2011). Nonhistone Scm3 binds to AT-rich DNA to organize atypical centromeric nucleosome of budding yeast. Mol. Cell *43*, 369–380.

Xu, J., Yanagisawa, Y., Tsankov, A.M., Hart, C., Aoki, K., Kommajosyula, N., Steinmann, K.E., Bochicchio, J., Russ, C., Regev, A., et al. (2012). Genome-wide identification and characterization of replication origins by deep sequencing. Genome Biol. *13*, R27.

Xu, M., Long, C., Chen, X., Huang, C., Chen, S., and Zhu, B. (2010). Partitioning of histone H3-H4 tetramers during DNA replication-dependent chromatin assembly. Science *328*, 94–98.

Xu, W., Aparicio, J.G., Aparicio, O.M., and Tavaré, S. (2006). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae. BMC Genomics *7*, 276.

Yan, Z., Delannoy, M., Ling, C., Daee, D., Osman, F., Muniandy, P.A., Shen, X., Oostra, A.B., Du, H., Steltenpool, J., et al. (2010). A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. Mol. Cell *37*, 865–878.

Yang, B., Britton, J., and Kirchmaier, A.L. (2008). Insights into the impact of histone acetylation and methylation on Sir protein recruitment, spreading, and silencing in Saccharomyces cerevisiae. J. Mol. Biol. *381*, 826–844.

Yang, C.H., Tomkiel, J., Saitoh, H., Johnson, D.H., and Earnshaw, W.C. (1996). Identification of overlapping DNA-binding and centromere-targeting domains in the human kinetochore protein CENP-C. Mol. Cell. Biol. *16*, 3576–3586.

Yoda, K., Kitagawa, K., Masumoto, H., Muro, Y., and Okazaki, T. (1992). A human centromere protein, CENP-B, has a DNA binding domain containing four potential alpha helices at the NH2 terminus, which is separable from dimerizing activity. J. Cell Biol. *119*, 1413–1427.

Yoda, K., Ando, S., Okuda, A., Kikuchi, A., and Okazaki, T. (1998). In vitro assembly of the CENP-B/alpha-satellite DNA/core histone complex: CENP-B causes nucleosome positioning. Genes to Cells *3*, 533–548.

Yoda, K., Ando, S., Morishita, S., Houmura, K., Hashimoto, K., Takeyasu, K., and Okazaki, T. (2000). Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution in vitro. Proc. Natl. Acad. Sci. U. S. A. *97*, 7266–7271.

Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626–630.

Zhang, K., Mosch, K., Fischle, W., and Grewal, S.I.S. (2008a). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat. Struct. Mol. Biol. *15*, 381–388.

Zhang, W., Colmenares, S.U., and Karpen, G.H. (2012). Assembly of Drosophila centromeric nucleosomes requires CID dimerization. Mol. Cell *45*, 263–269.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008b). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat. Struct. Mol. Biol. *16*, 847–852.

Zhang, Z., Hayashi, M.K., Merkel, O., Stillman, B., and Xu, R.-M. (2002). Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. EMBO J. *21*, 4600–4611.

Zhang, Z., Wippo, C.J., Wal, M., Ward, E., Korber, P., and Pugh, B.F. (2011). A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. Science *332*, 977–980.

Zinkowski, R.P., Meyne, J., and Brinkley, B.R. (1991). The centromere-kinetochore complex: a repeat subunit model. J. Cell Biol. *113*, 1091–1110.

Zisimopoulou, P., Staib, C., Nanda, I., Schmid, M., and Grummt, F. (1998). Mouse homolog of the yeast origin recognition complex subunit ORC1 and chromosomal localization of the cognate mouse gene Orc1. Mol. Gen. Genet. *260*, 295–299.

open Biology

rsob.royalsocietypublishing.org

Research

Cite this article: Subramanian L, Toda NRT, Rappsilber J, Allshire RC. 2014 Eic1 links Mis18 with the CCAN/Mis6/Ctf19 complex to promote CENP-A assembly. *Open Biol.* **4**: 140043. http://dx.doi.org/10.1098/rsob.140043

Received: 8 March 2014
Accepted: 3 April 2014

Subject Area:
biochemistry/cellular biology/genetics/molecular biology

Keywords:
CENP-A, centromeres, epigenetics, fission yeast, Mis18

Author for correspondence:
Robin C. Allshire
e-mail: robin.allshire@ed.ac.uk

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsob.140043.

Royal Society Publishing

Eic1 links Mis18 with the CCAN/Mis6/Ctf19 complex to promote CENP-A assembly

Lakxmi Subramanian[1], Nicholas R. T. Toda[1], Juri Rappsilber[1,2] and Robin C. Allshire[1]

[1]Wellcome Trust Centre for Cell Biology, Institute of Cell Biology, School of Biological Sciences, The University of Edinburgh, Edinburgh EH9 3JR, UK
[2]Institute of Bioanalytics, Department of Biotechnology, Technische Universität Berlin, 13353 Berlin, Germany

1. Summary

CENP-A chromatin forms the foundation for kinetochore assembly. Replication-independent incorporation of CENP-A at centromeres depends on its chaperone HJURP^{Scm3}, and Mis18 in vertebrates and fission yeast. The recruitment of Mis18 and HJURP^{Scm3} to centromeres is cell cycle regulated. Vertebrate Mis18 associates with Mis18BP1^{KNL2}, which is critical for the recruitment of Mis18 and HJURP^{Scm3}. We identify two novel fission yeast Mis18-interacting proteins (Eic1 and Eic2), components of the Mis18 complex. Eic1 is essential to maintain Cnp1^{CENP-A} at centromeres and is crucial for kinetochore integrity; Eic2 is dispensable. Eic1 also associates with Fta7$^{CENP-Q/Okp1}$, Cnl2^{Nkp2} and Mal2$^{CENP-O/Mcm21}$, components of the constitutive CCAN/Mis6/Ctf19 complex. No Mis18BP1^{KNL2} orthologue has been identified in fission yeast, consequently it remains unknown how the key Cnp1^{CENP-A} loading factor Mis18 is recruited. Our findings suggest that Eic1 serves a function analogous to that of Mis18BP1^{KNL2}, thus representing the functional counterpart of Mis18BP1^{KNL2} in fission yeast that connects with a module within the CCAN/Mis6/Ctf19 complex to allow the temporally regulated recruitment of the Mis18/Scm3^{HJURP} Cnp1^{CENP-A} loading factors. The novel interactions identified between CENP-A loading factors and the CCAN/Mis6/Ctf19 complex are likely to also contribute to CENP-A maintenance in other organisms.

2. Introduction

Genome integrity and stability are key to the propagation of genetic information through generations. Centromeres are the specialized sites on eukaryotic chromosomes where kinetochores, which govern spindle microtubule attachment, assemble during cell division. Centromere dysfunction, and consequent kinetochore instability, can therefore cause mis-segregation of chromosomes resulting in aneuploidy and genome instability. A plethora of evidence suggests that DNA sequence is neither necessary nor sufficient for the establishment of a functional centromere [1,2], and it is generally accepted that centromere specification is epigenetically regulated in many organisms [3,4].

CENP-A, a histone H3 variant, replaces canonical H3 in specialized nucleosomes that mark active centromeres. CENP-A is both necessary and sufficient to mediate the establishment and maintenance of functional kinetochores. CENP-A associates with a distinct set of proteins that allow it to promote its own propagation to preserve centromere location [5–7]. In the fission yeast *Schizosaccharomyces pombe*, Cnp1^{CENP-A} is restricted to the non-repetitive central domain of centromeres (approx. 7–10 kb; unique central core plus inverted inner-most repeats)

that is flanked by heterochromatic outer repeats. Cnp1^{CENP-A} recruitment to the central domain depends on the conserved kinetochore proteins Mis18, Mis6$^{CENP-I/Ctf3}$, Sim4$^{CENP-K/Mcm22}$ and Scm3^{HJURP} [8–12]. Mis18 was originally identified as being required to maintain Cnp1^{CENP-A} at fission yeast centromeres. Subsequently, Mis18 function was shown to be conserved at vertebrate centromeres [11,13]. Mis18 associates with the nuclear protein Mis16, which is orthologous to *Saccharomyces cerevisiae* Hat2 and human RbAp46/48, which are known to act as histone chaperones in many distinct histone transactions [14,15]. Mis6$^{CENP-I/Ctf3}$ and Sim4$^{CENP-K/Mcm22}$ reside in a larger complex consisting of approximately 16–18 centromeric proteins [16–18], most of which are conserved and collectively form the complex known as CCAN, Ctf19 and Mis6 in vertebrates, budding yeast and fission yeast, respectively [16,19,20]. The modular CCAN/Mis6/Ctf19 complex forms the main scaffold of centromere proteins that is required to recruit outer kinetochore components in order to mediate and regulate kinetochore–microtubule attachments [21]. Scm3^{HJURP} is orthologous to the vertebrate CENP-A-specific chaperone HJURP [22]. Scm3^{HJURP} directly associates with Cnp1^{CENP-A} and is required for Cnp1^{CENP-A} incorporation at centromeres [9,12].

In vertebrate cells, the assembly of new CENP-A into centromeric chromatin is uncoupled from replication and occurs in early G1, prior to S phase [23]. Consequently, following centromere replication, the level of resident CENP-A at centromeres is halved as a result of its distribution to the resulting two sister-centromeres [24]. The CENP-A loading factors Mis18α/β and Mis18BP1^{KNL2} and the CENP-A chaperone HJURP^{Scm3} are transiently recruited to human centromeres in telophase and are required to replenish CENP-A in G1 [13,25,26]. Mis18 is recruited to human centromeres by Mis18BP1^{KNL2}, which associates with centromeres via the C-terminus of the constitutive kinetochore component, CENP-C$^{Cnp3/Mif2}$ [27,28]. Both Mis18α and Mis18BP1^{KNL2} are required for the cell-cycle-regulated recruitment of HJURP^{Scm3} prior to G1 [25,28,29]. Interestingly, in fission yeast, the CENP-C orthologue Cnp3 is not essential for viability [30] and no Mis18BP1^{KNL2} protein can be identified. It therefore remains unclear how Mis16$^{RbAp46/48/Hat2}$, Mis18 and Scm3^{HJURP} are recruited to centromeres to maintain Cnp1^{CENP-A} at fission yeast centromeres. Nevertheless, as in vertebrate cells, the Cnp1^{CENP-A} assembly factors Mis16$^{RbAp46/48/Hat2}$, Mis18 and Scm3^{HJURP} all display dynamic association with fission yeast centromeres during the cell cycle, but they are released from centromeres in mitotic prophase and re-associate in mid-anaphase [9,11–13]. This suggests that fission yeast relies on an alternative to the CENP-C$^{Cnp3/Mif2}$–Mis18BP1^{KNL2} interaction for the regulated recruitment of the Mis16$^{RbAp46/48/Hat2}$, Mis18 and Scm3^{HJURP} Cnp1^{CENP-A} assembly factors to centromeres.

Although many kinetochore proteins have clearly been conserved between different experimental organisms over evolutionary time, it is also evident that particular proteins have been lost in some organisms and replaced by alternative pathways. For example, both *Drosophila* and *S. cerevisiae* lack the Mis18 and Mis18BP1^{KNL2} proteins. However, *S. cerevisiae* retains an Scm3^{HJURP} orthologue, whereas Scm3^{HJURP} function is replaced by the distinct Cal1 protein in *Drosophila* [31,32]. Moreover, in *Arabidopsis* the association dynamics of Mis18BP1^{KNL2} with centromeres is more similar to that of *S. pombe* Mis18/Scm3^{HJURP} than human Mis18BP1^{KNL2}/Mis18/HJURP^{Scm3} in that it departs only briefly from centromeres during mitosis, re-associating in mid-anaphase prior to *Arabidopsis* CenH3^{CENP-A} deposition in G2 [33,34]. It is well recognized that alternative solutions can evolve to mediate the same conserved process, and such alternatives can inform on parallel or other underlying pathways and mechanisms that may be more prevalent in one system relative to another.

In fission yeast, Mis16$^{RbAp46/48/Hat2}$–Mis18 may promote Cnp1^{CENP-A} assembly through a direct physical interaction detected *in vitro* with the Cnp1^{CENP-A} chaperone Scm3^{HJURP} [9,12]. However, as no direct interaction has been reported between Mis16$^{RbAp46/48/Hat2}$–Mis18 and constitutive CCAN components, it remains unknown how the Mis16$^{RbAp46/48/Hat2}$–Mis18 and Scm3^{HJURP} assembly factors are recruited to the underlying constitutive kinetochore scaffold. Here, we set out to determine how fission yeast solves the problem of recruiting the Cnp1^{CENP-A} loading machinery (Mis16$^{RbAp46/48/Hat2}$/Mis18/Scm3^{HJURP}) to centromeres without a Mis18BP1^{KNL2} orthologue. Through a proteomics approach, we identified two previously uncharacterized proteins as additional integral components of the Mis16$^{RbAp46/48/Hat2}$-Mis18 complex, designated Eic1 and Eic2 (Eighteen Interacting Centromere proteins 1 and 2). Our analyses indicate that Eic1 is essential to maintain normal Cnp1^{CENP-A} levels at centromeres, whereas Eic2 is dispensable. Importantly, we also demonstrate that Eic1 provides a link between Mis16$^{RbAp46/48/Hat2}$-Mis18 and specific constitutive components of the kinetochore scaffold (CCAN/Mis6/Ctf19 complex components). The analyses presented identify three components within the conserved CCAN/Mis6/Ctf19 complex which we propose represent a module that is required to recruit the Cnp1^{CENP-A} loading machinery (Mis16$^{RbAp46/48/Hat2}$/Mis18/Eic1/Scm3^{HJURP}) to fission yeast centromeres.

3. Results

3.1. Identification of novel proteins that associate with Mis16$^{RbAp46/48/Hat2}$ or Mis18

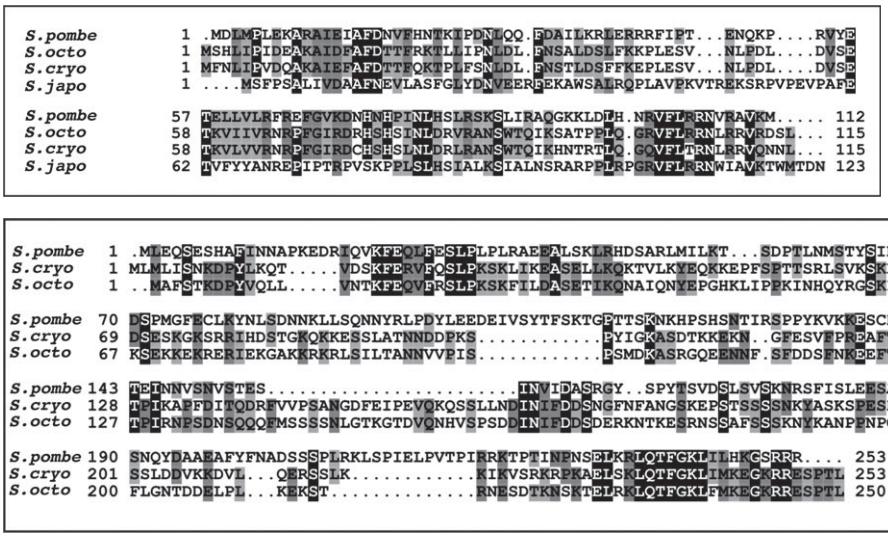
To explore the interactions of Mis16$^{RbAp46/48/Hat2}$ and Mis18 with other proteins in fission yeast, Myc-tagged Mis16 and Mis18 were immunoprecipitated from extracts of cells expressing either Mis16-Myc or Mis18-Myc as fusion proteins from the endogenous genes and the captured proteins subjected to LC-MS/MS analyses. Mis18 immunoprecipitates were reproducibly found to contain two previously uncharacterized proteins, which we named Eic1 (SPBC27B12.02) and Eic2 (SPBC776.16) for Eighteen Interacting Centromere protein 1 and 2, respectively. Mis16$^{RbAp46/48/Hat2}$ immunoprecipitates also reproducibly contained Eic1, along with two other known interactors: Mis18 and the histone H4 acetyltransferase Hat1 (figure 1*a*) [11,35]. Proteins orthologous to Eic1 were identifiable in all the sequenced genomes of the three other fission yeast species (*S. octosporus*, *S. cryophilus* and *S. japonicus*) by homology and synteny searches (figure 1*b*). However, no orthologue of Eic2 was apparent in *S. japonicus* (figure 1*b*).

Co-immunoprecipitation experiments verified the interactions of Eic1 and Eic2 with Mis16$^{RbAp46/48/Hat2}$ and Mis18. Consistent with the LC-MS/MS results, both Mis16-Myc and Mis18-Myc were found to co-immunoprecipitate with Eic1-GFP. Also, Mis18-Myc, but not Mis16-Myc, co-immunoprecipitated with Eic2-GFP (figure 1*c*). This demonstrates that Mis16$^{RbAp46/48/Hat2}$, Mis18, Eic1 and Eic2 associate.

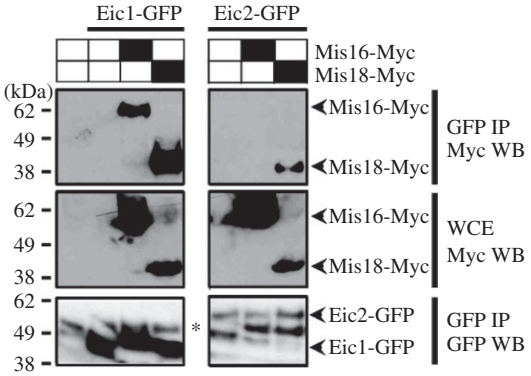
(*a*)

protein	mol. weight (kDa)	Mis16$^{RbAp46/48/Hat2}$-Myc IP (avg. no. unique peptides)	Mis18-Myc IP (avg. no. unique peptides)
Mis16	48.43	3.667	—
Hat1	44.05	8.333	—
Mis18	22.54	4.333	7.333
Eic1 (SPBC27B12.02)	13.43	3.667	3.000
Eic2 (SPBC776.16)	28.73	—	1.667

(*b*)



(*c*)



(*d*)

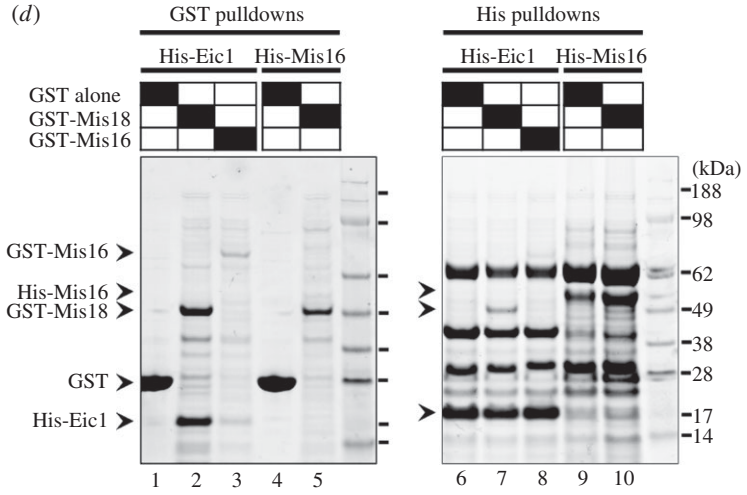


Figure 1. Eic1 and Eic2 are Mis18-interacting proteins. (*a*) LC-MS/MS analysis of Myc-tagged Mis16$^{RbAp46/48/Hat2}$ and Mis18 immunoprecipitates from *S. pombe* whole cell extracts identifies two previously uncharacterized proteins, Eic1 (SPBC27B12.02) and Eic2 (SPBC776.16). Average number of unique peptides reproducibly identified from three independent experiments is shown. (*b*) Primary sequence alignments of Eic1 (top) and Eic2 (bottom) orthologues identified among *Schizosaccharomyces* species. (*c*) Eic1-GFP co-immunoprecipitates with both Mis16-Myc and Mis18-Myc, while Eic2-GFP only co-immunoprecipitates with Mis18-Myc. The asterisk (*) in the bottom panel denotes the IgG heavy chain. (*d*) Eic1 directly interacts with Mis18 and Mis16. 6xHis-Eic1 was co-expressed with GST alone, GST-Mis18 or GST-Mis16 in *E. coli*. Coomassie-stained SDS-PAGE gels showing reciprocal GST (lanes 1–3) and His (lanes 6–8) pulldowns from *E. coli* lysates are shown. Also shown are reciprocal pulldowns of 6xHis-Mis16 co-expressed with GST alone or GST-Mis18 (lanes 4–5 and 9–10).

To determine whether Eic1 can directly associate with Mis16$^{RbAp46/48/Hat2}$ or Mis18, we co-expressed 6xHis-Eic1 with GST-Mis16 or GST-Mis18 in *Escherichia coli*. GST-Mis18 and 6xHis-Eic1 can clearly associate with each other in reciprocal pull down assays (figure 1*d*, lanes 2 and 7). 6xHis-Eic1 could also be detected in pulldowns of GST-Mis16 (figure 1*d*, lane 3). Similarly, a weak interaction between 6xHis-Mis16 and GST-Mis18 could be detected *in vitro* (figure 1*d*, lane 5 when compared with lanes 4 or 2).

3.2. Hat1 does not contribute to the maintenance of Cnp1^{CENP-A} chromatin

In *S. cerevisiae*, Hat2 is the orthologue of Mis16 and it has previously been shown to associate with the histone H4 acetyl-transferase Hat1 [36]. Related to this, levels of acetylated histones have been shown to increase at centromeres in *mis16* and *mis18* mutants [11]. It is possible that altered histone acetylation at centromeres relates to the function of

Mis16$^{RbAp46/48/Hat2}$ and Hat1 at centromeres. We first confirmed that Mis16-Myc and Hat1-HA co-immunoprecipitate (electronic supplementary material, figure S1*a*) and then, to investigate the possibility that Hat1 possesses centromere-specific functions, we subjected Hat1-HA immunoprecipitates to LC/MS-MS analyses. However, we were unable to detect any Hat1-associated proteins apart from Mis16$^{RbAp46/48/Hat2}$, confirming the analyses of others [35]. Moreover, Hat1-HA was not found to be enriched in the central kinetochore domain of centromeres (electronic supplementary material, figure S1*b*) and no loss of Cnp1^{CENP-A} from centromeres could be detected in cells lacking Hat1 (*hat1*Δ; electronic supplementary material, figure S1*c*). Thus, it appears that Mis16$^{RbAp46/48/Hat2}$ may participate in two distinct functional complexes: Mis16$^{RbAp46/48/Hat2}$–Mis18–Eic1–Eic2 at centromeres (see below) and Mis16$^{RbAp46/48/Hat2}$–Hat1 elsewhere in the nucleus (electronic supplementary material, figure S1*d*). Such a multi-functional role for Mis16$^{RbAp46/48/Hat2}$ would be consistent with the known involvement of RbAp46/48 proteins in many complexes involved in histone modification and chromatin remodelling [14,15].

3.3. Eic1 and Eic2 associate with centromeres dynamically through the cell cycle

CENP-A and all known fission yeast kinetochore proteins localize specifically at centromeres and are enriched over the central domain region of centromeres. Both Eic1 and Eic2 associate with Mis18, which is known to localize to fission yeast centromeres for most of the cell cycle, apart from early prophase to mid-anaphase of mitosis [13]. To examine the localization of Eic1 and Eic2, the endogenous genes expressing Eic1 and Eic2 were fused to GFP. Quantitative chromatin immunoprecipitation (qChIP) analyses demonstrated that, as with Mis18 and other kinetochore proteins, both Eic1-GFP and Eic2-GFP are enriched over the central kinetochore domain (central core plus imr repeats) of fission yeast centromeres (figure 2*a*). Genome-wide ChIP-seq analyses confirmed this finding and also demonstrated that Eic1-GFP and Eic2-GFP, much like Mis18-GFP and Scm3-GFP, are undetectable at other regions of the genome including the heterochromatic outer repeats of centromeres (figure 2*b* and electronic supplementary material, figure S2). Immunostaining revealed that both Eic1 and Eic2 GFP-tagged proteins co-localize with Cnp1^{CENP-A} at clustered centromeres in interphase cells (figure 2*c*(i),(ii)). However, as with Mis16$^{RbAp46/48/Hat2}$ and Mis18 [11,13], Eic1-GFP and Eic2-GFP dissociate from centromeres in early mitosis (prometaphase), just as the centromeres (Cnp1) begin to form two separate clusters as they biorient on the spindle (figure 2*c*(iii),(iv)). Moreover, both Eic1-GFP and Eic2-GFP reassociate with centromeres in mid-anaphase when centromeres (Cnp1) and chromosomes (DAPI) have clearly segregated to opposite spindle poles (figure 2*c*(viii),(ix)). Apart from Mis16$^{RbAp46/48/Hat2}$ and Mis18, the CENP-A chaperone Scm3^{HJURP} exhibits similar temporal dissociation–reassociation at centromeres during mitosis [9,12]. Thus, our analyses identify Eic1 and Eic2 as two additional proteins that are released from centromeres in early mitosis and reloaded on centromeres in mid-anaphase. The finding that Eic1 and Eic2 exhibit similar association dynamics as Mis18 through the cell cycle is consistent with our identification of Eic1 and Eic2 being tightly associated with Mis18.

3.4. Eic1 is required to maintain Cnp1^{CENP-A} at centromeres

Mis18 and associated proteins have been shown to be required to maintain CENP-A at centromeres in fission yeast and vertebrates [11,13,37]. If Eic1 and Eic2 are critical for Mis18 function then they should also be required to maintain normal Cnp1^{CENP-A} levels at centromeres. As the *eic1*$^{+}$ gene is essential for cell viability, we generated a conditional temperature-sensitive (*ts*) mutant of *eic1* (*eic1-1*:*hygR*, hereafter referred to as *eic1-1*) in which a single amino acid substitution (Phe102Ser) at a conserved residue rendered cells inviable at 36°C but with retained viability at 25°C (figure 3*a*). We also generated a corresponding wild-type allele of *eic1* (*eic1*$^{+}$:*hyg*^{R}) as a control in which, as with *eic1-1*, a hygromycin resistance marker was inserted within the 3′UTR of *eic1* at its endogenous locus for ease of genetic manipulation (figure 3*a*). qChIP analyses demonstrated that Cnp1^{CENP-A} levels were significantly diminished at centromeres in *eic1-1* cells at restrictive temperature (figure 3*b*; electronic supplementary material, figure S3*a*). Cells harbouring a second *ts* allele of *eic1* (*eic1-2*:*hyg*^{R}; two substituted residues, Lys37Glu and Tyr55His, hereafter referred to as *eic1-2*) also demonstrated significant loss of Cnp1^{CENP-A} from centromeres at 36°C (figure 3*a*,*b*; electronic supplementary material, figure S3*a*). Furthermore, *eic1-1* and *eic1-2* cells displayed sensitivity to the microtubule depolymerizing drug thiabendazole (TBZ), whereas *eic1*$^{+}$:*hyg*^{R} and the *mis18-262* ts mutant did not confer TBZ sensitivity (figure 3*c*). TBZ sensitivity suggests that kinetochore–microtubule interactions are defective in *eic1* mutants. Aberrant kinetochore function in *eic1-1* cells was supported by cytological analyses, which revealed severe chromosome segregation defects (electronic supplementary material, figure S3*b*). qChIP analyses demonstrated that GFP-tagged Eic1-1 and Eic1-2 mutant proteins remained associated with centromeres even at 36°C; thus, the observed phenotypes of *eic1-1* and *eic1-2* cells are not a consequence of the complete absence of Eic1 protein at centromeres (electronic supplementary material, figure S3*c*).

In contrast to *eic1* mutants, cells lacking the *eic2*$^{+}$ gene (*eic2*Δ) displayed no loss of Cnp1^{CENP-A} from centromeres (figure 3*d*). One possibility is that Eic2 is not required to maintain Cnp1^{CENP-A} at centromeres but is required to establish Cnp1^{CENP-A} on naive centromeric DNA. To test this, a plasmid bearing centromeric DNA that efficiently establishes Cnp1^{CENP-A} chromatin and functional kinetochores in wild-type cells (pH-cc2) [38] was transformed into *eic2*Δ cells. No defect in the establishment of Cnp1^{CENP-A} chromatin or functional kinetochores was observed (electronic supplementary material, figure S4). Furthermore, *eic2*Δ cells were not sensitive to TBZ (figure 3*c*). We conclude that Eic1 is required for Cnp1^{CENP-A} maintenance and kinetochore integrity, whereas Eic2 is dispensable.

3.5. The recruitment of Cnp1^{CENP-A} assembly factors is reduced at centromeres in *eic1* and *eic2* mutants

Previous studies have shown that defective Mis16$^{RbAp46/48/Hat2}$ or Mis18 function affects the localization of the Cnp1^{CENP-A} loading factors Scm3^{HJURP} and Mis6$^{CENP-I/Ctf3}$ at centromeres [9,11,12]. Vertebrate Mis18α/β and Mis18BP1^{KNL2} have also been shown to be required for the recruitment of the CENP-A

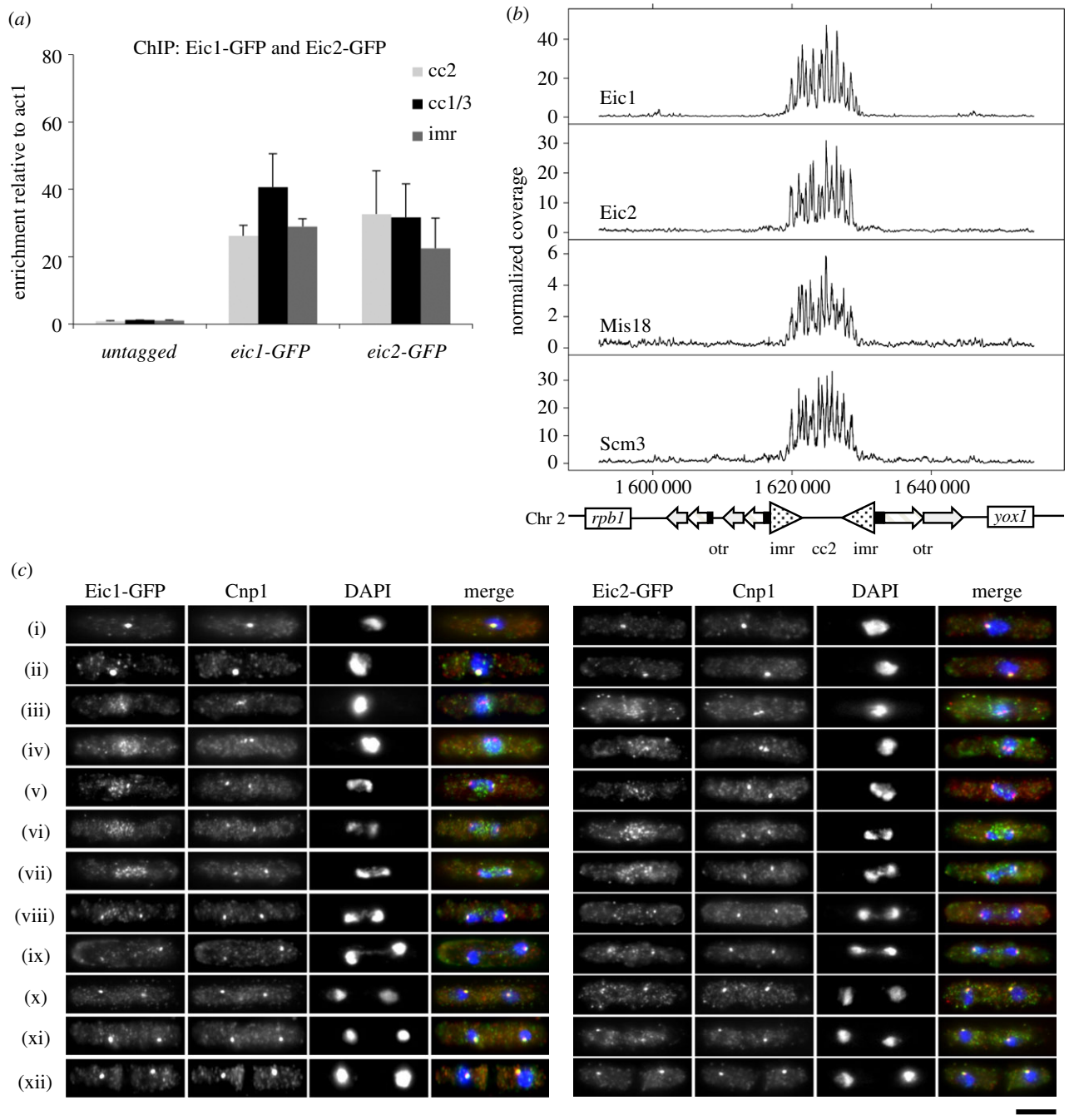


Figure 2. Eic1 and Eic2 associate specifically with centromeres. (*a*) Eic1 and Eic2 bind the central domain of *S. pombe* centromeres. qChIP analyses showing enrichments of GFP-tagged Eic1 and Eic2 at the central cores (cc) of centromeres 1, 2 and 3 and *imr* repeats of centromere 1, relative to the *act1* locus. Error bars represent standard deviation between at least three biological replicates. (*b*) Eic1 and Eic2 exhibit very similar genome-wide association profiles as Mis18 and Scm3^{HJURP}. A comparison between the ChIP-seq profiles of GFP-tagged Eic1, Eic2, Mis18 and Scm3 across centromere 2 is presented, alongside a schematic diagram of centromere 2 (bottom). Normalized coverage represents the number of sequencing fragments obtained from anti-GFP IP normalized to that obtained from the input. (*c*) Eic1 and Eic2 exhibit very similar cell-cycle localization dynamics as Mis18 and Scm3^{HJURP}. Immunofluorescence of *S. pombe* cells expressing GFP-tagged Eic1 or Eic2 stained with antibodies to GFP (green) and Cnp1^{CENP-A} (red), and DAPI (blue). Both Eic1 and Eic2 dissociate from centromeres during prometaphase to mid-anaphase of mitosis ((iii)–(vii)) and subsequently reassociate. Scale bar, 5 μm.

chaperone HJURP^{Scm3} to centromeres, and consequently the incorporation of CENP-A at centromeres [13,25]. We find that the levels of GFP-tagged Mis16$^{RbAp46/48/Hat2}$ and Mis18 associated with fission yeast centromeres are greatly reduced in the *eic1-1* mutant even at 25°C (figure 4*a*,*b*), consistent with the observed synthetic genetic interaction between *eic1-1* and the GFP-tagged alleles of *mis16* and *mis18* (electronic supplementary material, figure S5). Additionally, we find that Scm3^{HJURP} association with centromeres is also dependent on Eic1 function (figure 4*c*). Interestingly, we find Mis6$^{CENP-I/Ctf3}$ association with centromeres to be only partially dependent on Eic1 (figure 4*d*), as intermediate levels of Mis6$^{CENP-I/Ctf3}$ are retained at centromeres even when Eic1 function is compromised. Thus Eic1, in conjunction with its interacting partners Mis16$^{RbAp46/48/Hat2}$ and Mis18, is essential to maintain normal levels of Cnp1^{CENP-A} as well as Cnp1^{CENP-A} loading factors on centromeres.

Surprisingly, *eic2*Δ cells exhibited reduced levels of Mis18, Mis16$^{RbAp46/48/Hat2}$ and Mis6$^{CENP-I/Ctf3}$ at centromeres, whereas Scm3^{HJURP} remained unaffected (figure 4*e*–*h*). As normal levels of Cnp1^{CENP-A} are retained at centromeres and no defect in chromosome segregation was detectable in *eic2*Δ cells

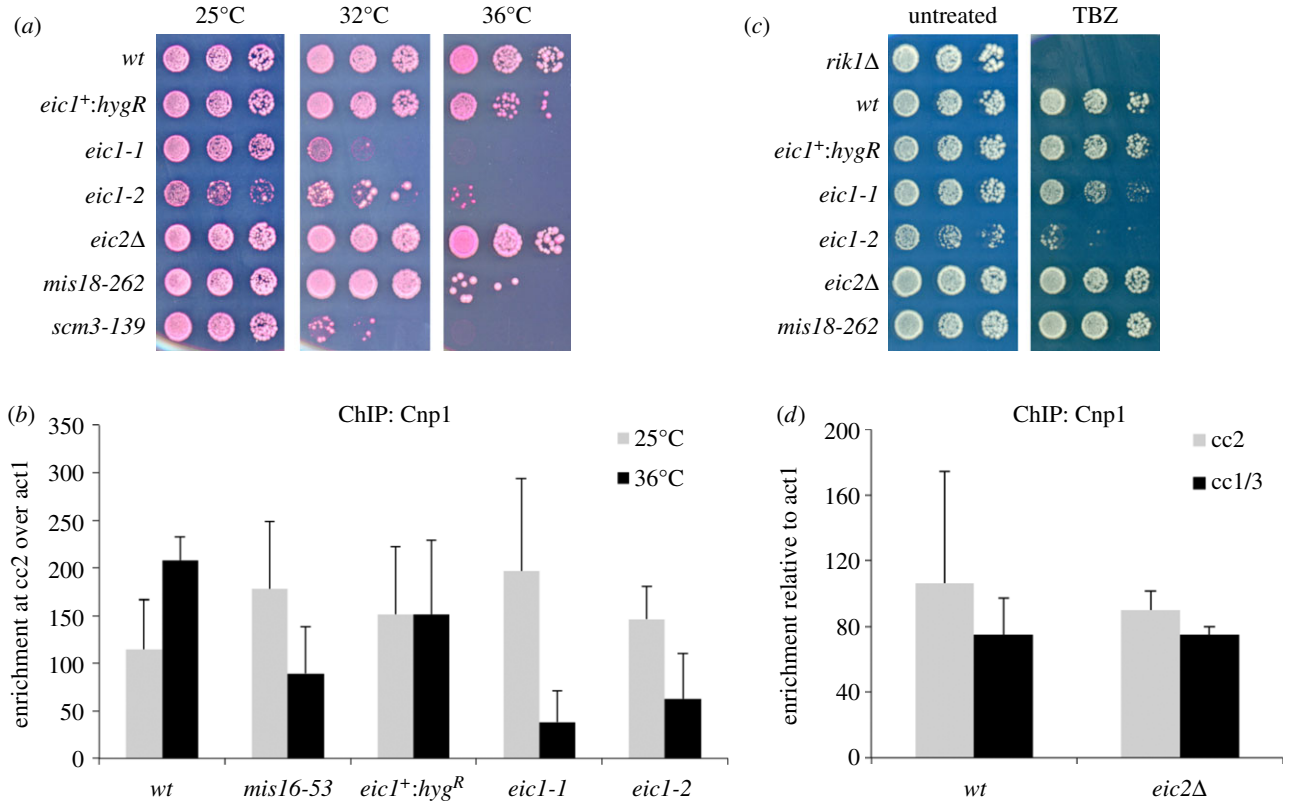


Figure 3. Eic1 is required for Cnp1^{CENP-A} assembly, while Eic2 is dispensable. (*a*) *ts* mutations in *eic1* affect cell viability, while *eic2*Δ cells show no defects in growth. Five-fold serial dilutions of cells spotted on YES + Phloxine B media and incubated at the indicated temperatures; dead cells stain dark pink. (*b*) *eic1* mutants display reduced Cnp1^{CENP-A} levels at centromeres. qChIP analyses of Cnp1^{CENP-A} association with centromeres in the indicated strains when grown at permissive (25°C) versus restrictive temperature (36°C) for 8 h. Enrichment of cc2 DNA relative to the *act1* locus is presented. (*c*) *eic1* mutants display sensitivity to TBZ, while *eic2*Δ cells show no TBZ sensitivity. Five-fold serial dilutions of cells spotted on YES media (untreated) or YES media supplemented with 12.5 μg ml^{-1} TBZ, and incubated at 25°C. (*d*) *eic2*Δ cells display no loss of Cnp1^{CENP-A} at centromeres. qChIP analyses of Cnp1^{CENP-A} association with centromeres in the indicated strains when grown at 32°C. Enrichment of cc2 or cc1/3 DNA relative to the *act1* locus is presented. Error bars in (*b*,*d*) represent standard deviation between at least three biological replicates.

(figure 3*c*,*d* and electronic supplementary material, figure S4), the observed reduction of Mis18, Mis16$^{RbAp46/48/Hat2}$ and Mis6$^{CENP-I/Ctf3}$ at centromeres must not be sufficient to significantly affect Cnp1^{CENP-A} maintenance.

Epistasis analyses (double mutant combinations) frequently reveal positive and negative genetic interactions between mutants and consequently inform on the functional niche of specific proteins. The *eic1-1* mutation exhibited significantly reduced growth when combined with *ts* mutations in *mis18*, *scm3*, *cnp1* and *mis6* (figure 5*a* and electronic supplementary material, figure S6*a*; summarized in figure 5*c*). We failed to generate *eic1-1 mis16-53* and *eic1-1 scm3-139* double mutant strains, suggesting that *eic1-1* has a synthetic lethal interaction with both *mis16-53* and *scm3-139*. Surprisingly, such synthetic interactions could also be observed for the *eic1*$^{+}$*:hyg*^{R} allele, which must compromise *eic1* function in sensitized genetic backgrounds and therefore be considered a hypomorphic allele of *eic1* (figure 5 and electronic supplementary material, figure S6*a*). These results, in conjunction with our ChIP analyses of Cnp1^{CENP-A} assembly factors (figure 4*a*–*d*), clearly demonstrate that Eic1 makes important contributions to the proper localization of Cnp1^{CENP-A} assembly factors and Cnp1^{CENP-A} itself.

*eic2*Δ showed no combinatorial effects when combined with *cnp1*, *mis6* or *mis16 ts* mutants (electronic supplementary material, figure S6*b*; summarized in figure 5*c*). However, *eic2*Δ displayed a negative interaction with *mis18-262*, notably reducing growth at 36°C relative to *mis18-262* alone (figure 5*b*). By contrast, *eic2*Δ partially suppressed the temperature sensitivity of the *scm3-139* (Leu73Phe) but not the *scm3-15* (Ser281Leu) mutant (figure 5*b*; summarized in figure 5*c*). The negative interaction of Eic2 with Mis18 confirms that Eic2 contributes to the function of the Mis18 complex. As the protein levels as well as association of Scm3^{HJURP} with centromeres remain unaffected in *eic2*Δ cells (figure 4*c*; electronic supplementary material, figure S6*c*), the mechanism of partial suppression of temperature sensitivity observed in *eic2*Δ *scm3-139* cells remains unclear. *eic2*Δ showed no combinatorial effects when combined with *eic1-1* (figure 5*b*; summarized in figure 5*c*), suggesting that Eic2 may not influence Eic1 function even though Eic1 and Eic2 physically associate.

3.6. Eic1 and Eic2 association with centromeres is dependent on Cnp1^{CENP-A} loading factors

Previous analyses in fission yeast have shown that Mis16$^{RbAp46/48/Hat2}$ and Mis18 are required for the localization of the Cnp1^{CENP-A}-specific chaperone Scm3^{HJURP} and the CCAN protein Mis6$^{CENP-I/Ctf3}$ to centromeres [9,11,12]. Moreover, the localization of Mis18 at centromeres is unaffected by mutations in Cnp1^{CENP-A}, Mis6$^{CENP-I/Ctf3}$ or Scm3^{HJURP}, whereas Scm3^{HJURP} is dependent on functional Mis6$^{CENP-I/Ctf3}$, Mis16$^{RbAp46/48/Hat2}$ and Mis18 for its centromeric localization [9,11,12]. Such analyses suggest that Mis18, and probably Mis16$^{RbAp46/48/Hat2}$, function to mediate the

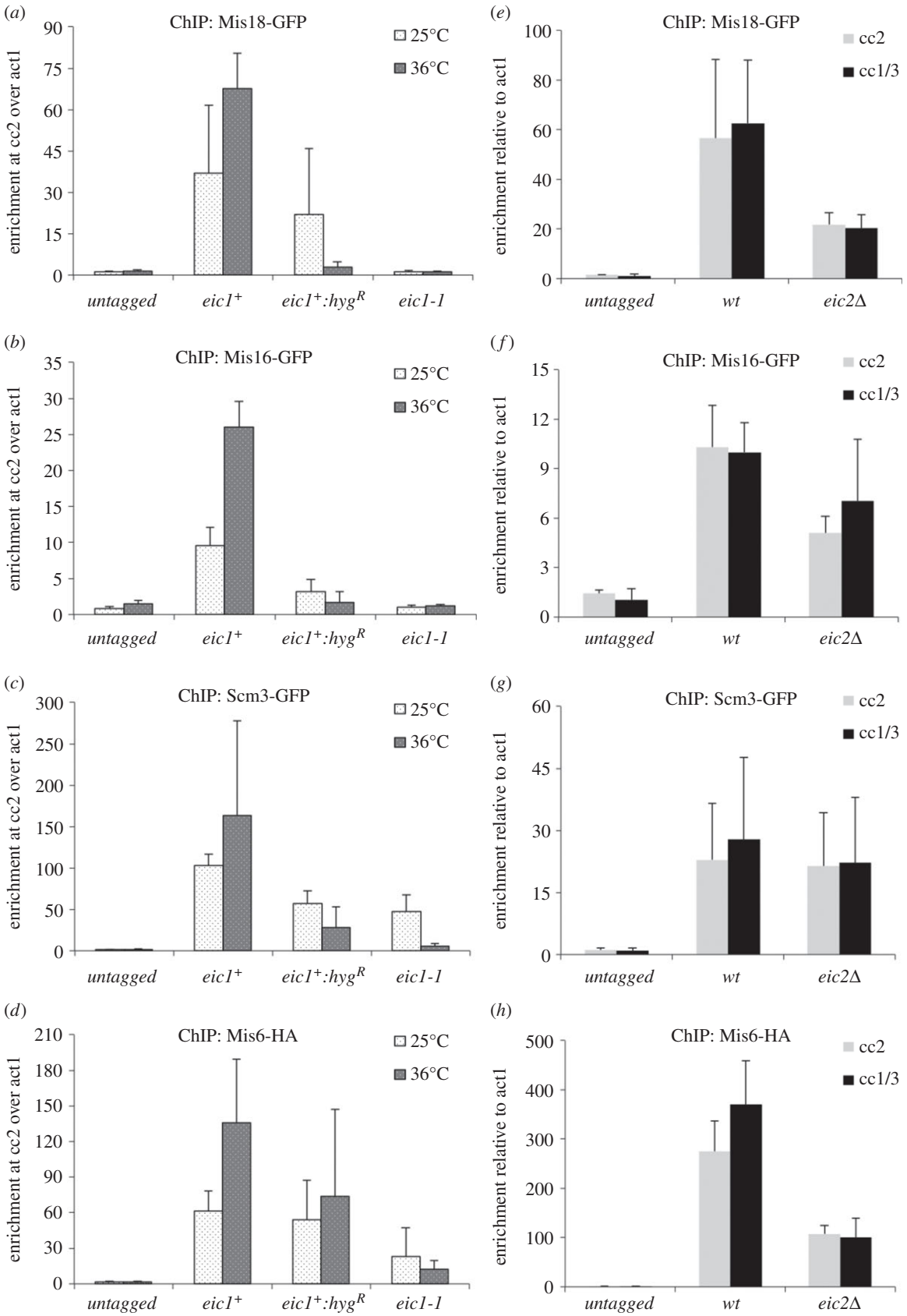
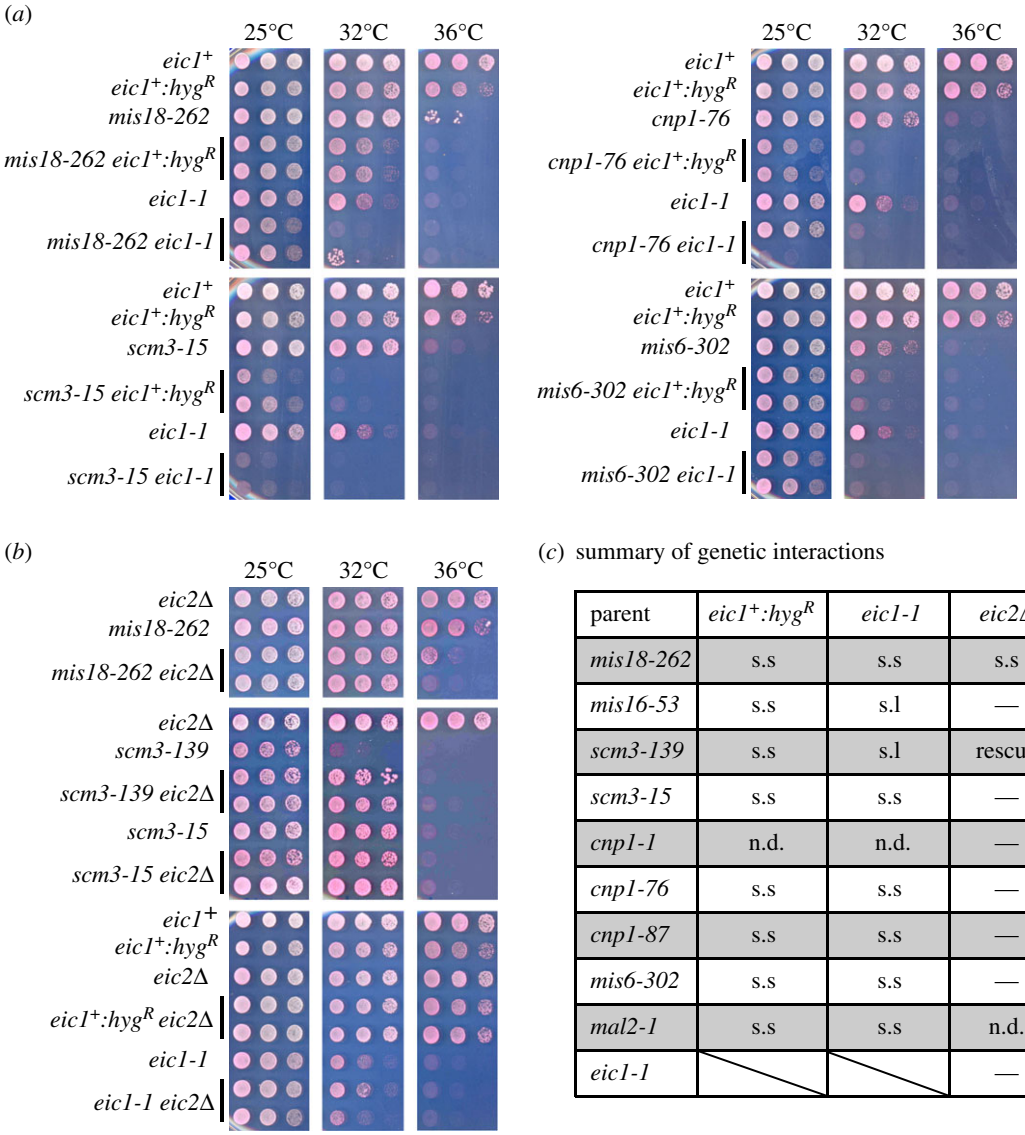


Figure 4. Eic1 and Eic2 promote normal levels of association of Cnp1^{CENP-A} assembly factors with centromeres. (*a*,*b*) Mis18-GFP and Mis16-GFP association with centromeres is entirely dependent on Eic1, and (*e*,*f*) partly dependent on Eic2. (*c*) Scm3-GFP association with centromeres is dependent on Eic1, and (*g*) largely independent of Eic2. (*d*) Mis6-HA association with centromeres is partly dependent on Eic1, and (*h*) Eic2. qChIP analyses of (*a*,*e*) Mis18-GFP, (*b*,*f*) Mis16-GFP, (*c*,*g*) Scm3-GFP and (*d*,*h*) Mis6-HA association with centromeres in the indicated strains when grown at permissive (25°C) versus restrictive temperature (36°C) for 8 h (*a*–*d*); or at 32°C (*e*–*h*). Enrichment of cc2 DNA relative to the *act1* locus is presented in (*a*–*d*). Enrichment of cc2 or cc1/3 DNA relative to the *act1* locus is presented in (*e*–*h*). Error bars represent standard deviation between at least three biological replicates.

recruitment of the Cnp1^{CENP-A} assembly factors Mis6$^{CENP-I/Ctf3}$ and Scm3^{HJURP} to centromeres and thus the incorporation of Cnp1^{CENP-A} itself. To further dissect the role of Eic1 and Eic2, and to determine whether they act together with Mis16$^{RbAp46/48/Hat2}$ and Mis18, we used qChIP to examine the dependencies of Eic1 and Eic2 for centromere localization



(c) summary of genetic interactions

parent	*eic1*$^{+}$*:hyg*^{R}	*eic1-1*	*eic2Δ*
mis18-262	s.s	s.s	s.s
mis16-53	s.s	s.l	—
scm3-139	s.s	s.l	rescue
scm3-15	s.s	s.s	—
cnp1-1	n.d.	n.d.	—
cnp1-76	s.s	s.s	—
cnp1-87	s.s	s.s	—
mis6-302	s.s	s.s	—
mal2-1	s.s	s.s	n.d.
eic1-1			—

Figure 5. Analysis of genetic interactions between *eic1* or *eic2* mutants and mutations in Cnp1^{CENP-A} or Cnp1^{CENP-A} assembly factors. (*a*) *eic1*$^{+}$*:hyg*^{R} and *eic1-1* cells display reduced growth when combined with mutations in *mis18*, *scm3*, *cnp1* or *mis6*. (*b*) *eic2*Δ cells display genetic interactions when combined with *mis18-262* and *scm3-139*, but not *eic1-1*. Five-fold serial dilutions of cells spotted on YES + Phloxine B media and incubated at the indicated temperatures; dead cells stain dark pink. (*c*) A tabular summary of genetic interactions analysed in this study. s.s, synthetic sick/reduced growth; s.l, synthetic lethal; n.d., not determined; —, no interaction detected.

in strains expressing various *ts* kinetochore proteins. Eic1 association with centromeres was found to be entirely dependent on functional Mis18, as indicated by undetectable levels of Eic1 at centromeres in *mis18-262* cells grown at 36°C (figure 6*a* and electronic supplementary material, figure S7*a*). Lower Eic1 levels were also detected on centromeric central cores in *mis16-53* and *scm3-139* mutants at restrictive temperature; however, they were largely unchanged in cells expressing mutant Cnp1^{CENP-A}, Mis6$^{CENP-I/Ctf3}$ and Mis12 proteins (figure 6*a* and electronic supplementary material, figure S7*a*). Centromeric Eic1 levels also remained unaffected in *eic2*Δ cells (figure 6*b*). Thus, the localization of Eic1 at kinetochores is mainly dependent on its partner proteins Mis16$^{RbAp46/48/Hat2}$ and Mis18, but not Eic2.

The level of Eic2 at centromeres was found to be greatly dependent on functional Mis16$^{RbAp46/48/Hat2}$, Mis18 and Eic1, and partially dependent on functional Scm3^{HJURP}, Cnp1^{CENP-A} and Mis6$^{CENP-I/Ctf3}$ (figure 6*c*,*d* and electronic supplementary material, figure S7*b*). Unexpectedly, Eic2 levels at centromeres were found to be reduced in *mis12-537* (a mutation in an outer kinetochore component) cells at restrictive temperature (figure 6*c* and electronic supplementary material, figure S7*b*). These analyses suggest that Eic2 is also recruited to centromeres by other mechanisms that are independent of its association with Mis18.

3.7. Eic1 associates with CCAN/Mis6/Ctf19 complex components that influence its recruitment to centromeres

To further investigate the functional niche of Eic1 and Eic2, immunoprecipitates of both GFP-tagged proteins were subjected to LC-MS/MS analyses to identify associated factors. Only Eic1 and Mis18 were detected in Eic2-GFP immunoprecipitates (figure 7*a*). However, three CCAN/Mis6/Ctf19 complex components were found to associate with Eic1-GFP: Fta7$^{CENP-Q/Okp1}$, Cnl2^{Nkp2} and Mal2$^{CENP-O/Mcm21}$ (figure 7*a*). Co-immunoprecipitation of Myc-tagged Fta7$^{CENP-Q/Okp1}$ or Cnl2^{Nkp2} with Eic1-GFP verified that these

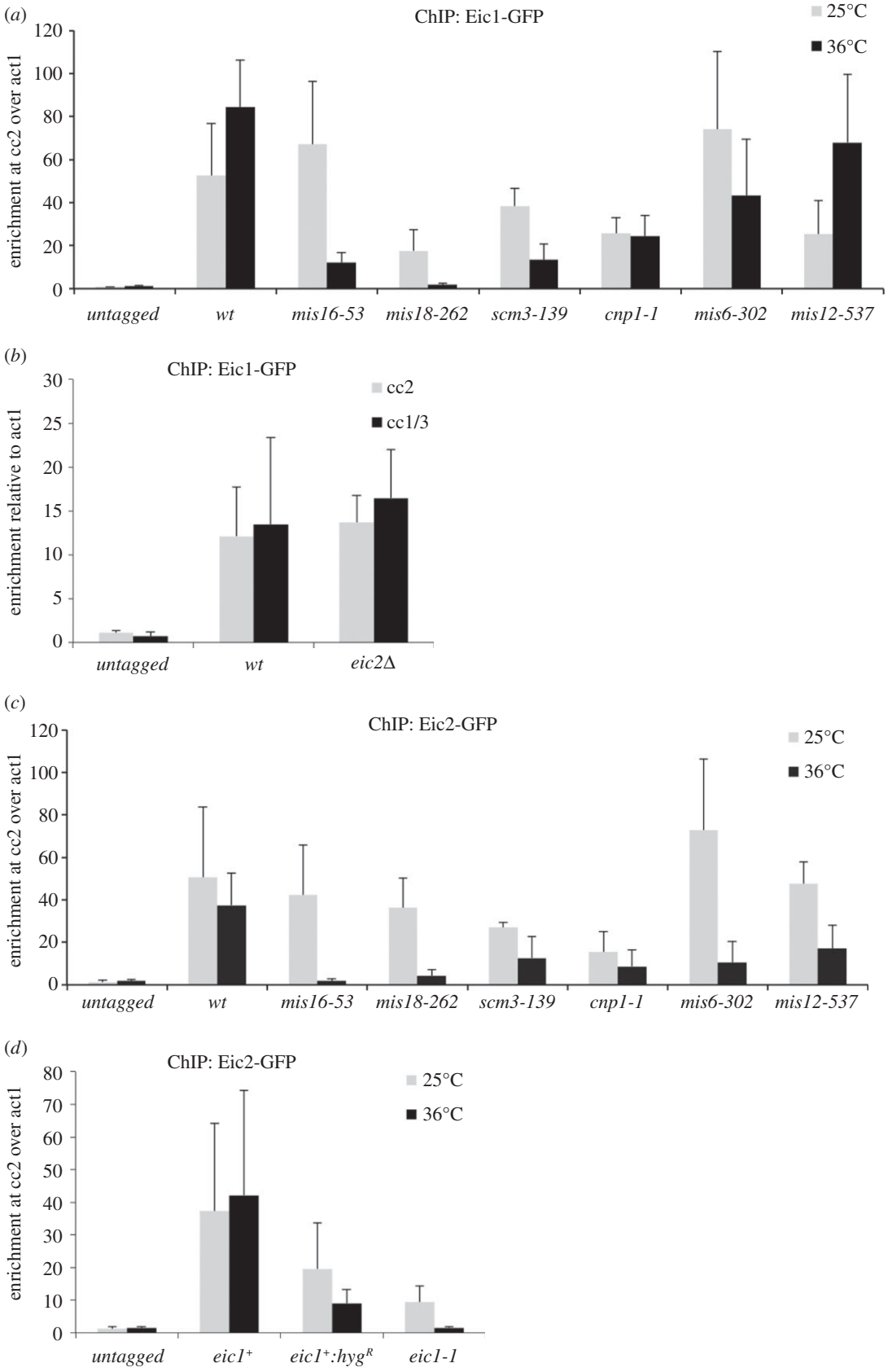


Figure 6. Eic1 and Eic2 depend on distinct Cnp1^{CENP-A} assembly factors for their association with centromeres. (*a*,*b*) Eic1 association with centromeres is dependent on Mis18, Mis16$^{RbAp46/48/Hat2}$ and Scm3^{HJURP}, but is largely independent of Cnp1^{CENP-A}, Mis6$^{CENP-I/Ctf3}$, Mis12 and Eic2. qChIP analyses of Eic1-GFP association with centromeres in the indicated strains when grown at permissive (25°C) versus restrictive temperature (36°C) for 8 h (*a*), or when grown at 32°C (*b*). (*c*,*d*) Eic2 association with centromeres is dependent on Mis18, Mis16$^{RbAp46/48/Hat2}$, Scm3^{HJURP}, Cnp1^{CENP-A}, Mis6$^{CENP-I/Ctf3}$, Mis12 and Eic1. qChIP analyses of Eic2-GFP association with centromeres in the indicated strains when grown at permissive (25°C) versus restrictive temperature (36°C) for 8 h. Enrichment of cc2 DNA relative to the *act1* locus is presented in (*a*,*c*,*d*). Enrichment of cc2 or cc1/3 DNA relative to the *act1* locus is presented in (*b*). Error bars represent standard deviation between at least three biological replicates.

(*a*)

protein	mol. weight (kDa)	Eic1-GFP IP (avg no. unique peptides)	Eic2-GFP IP (avg no. unique peptides)
Eic1	13.43	8.667	4.667
Eic2	28.73	16.667	12.000
Mis16	48.43	10.000	—
Mis18	22.54	11.333	7.667
Fta7$^{CENP-Q/Okp1}$	28.84	4.000	—
Cnl2^{Nkp2}	21.92	1.667	—
Mal2$^{CENP-O/Mcm21}$	34.52	2.667	—

(*b*)

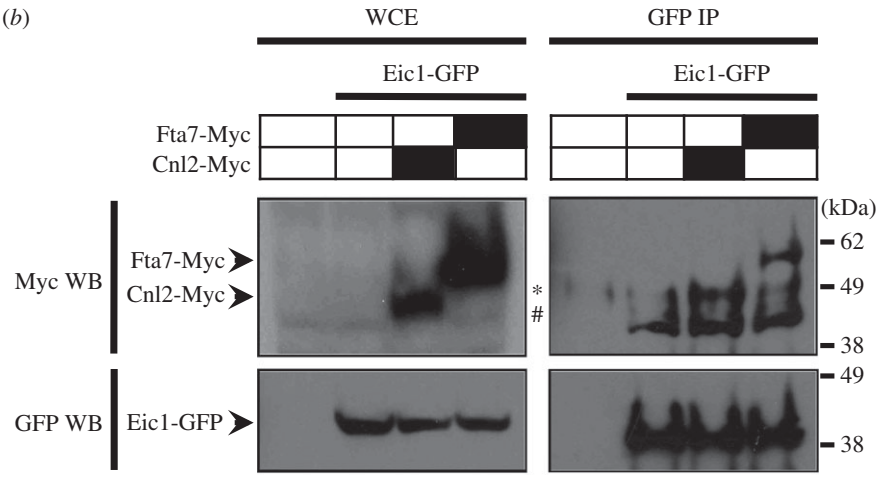


Figure 7. Eic1 interacts with three essential subunits of the CCAN/Mis6/Ctf19 complex. (*a*) LC-MS/MS analysis of GFP-tagged Eic1 or Eic2 immunoprecipitates from *S. pombe* whole cell extracts identifies Fta7$^{CENP-Q/Okp1}$, Cnl2^{Nkp2} and Mal2$^{CENP-O/Mcm21}$ as Eic1-interacting proteins. Average number of unique peptides reproducibly identified from three independent experiments is shown. (*b*) Eic1-GFP co-immunoprecipitates with Fta7-Myc and Cnl2-Myc. In the top panel, the asterisk (*) denotes the IgG heavy chain, and the hash tag (#) denotes a non-specific band.

CCAN/Mis6/Ctf19 complex components associate with Eic1 (figure 7*b*). ChIP analyses revealed that the association of Myc-tagged Fta7$^{CENP-Q/Okp1}$, Myc-tagged Cnl2^{Nkp2} and GFP-tagged Mal2$^{CENP-O/Mcm21}$ with centromeres is only partially dependent on functional Eic1 (figure 8*a–c*), as these proteins are detectable to a significant degree at centromeres in *eic1-1* cells even at non-permissive temperature. This is consistent with the finding that centromeric association of the CCAN component Mis6$^{CENP-I/Ctf3}$ also only partly depends on Eic1 function (figure 4*d*).

Interestingly, Eic1-GFP association with centromeres was found to greatly depend on functional Mal2$^{CENP-O/Mcm21}$ (figure 8*d*), although it appeared to be largely independent of functional Mis6$^{CENP-I/Ctf3}$ (figure 6*a* and electronic supplementary material, figure S7*a*). Additionally, the *mal2-1* ts mutant exhibited a more severe growth defect than *mis6-302* when combined with *eic1-1* (figures 5*a* and 8*e*). Together, these results suggest that the CCAN components Fta7$^{CENP-Q/Okp1}$, Cnl2^{Nkp2} and Mal2$^{CENP-O/Mcm21}$ may act in concert to recruit Eic1 to centromeres via their physical association with Eic1.

4. Discussion

In this study, we have identified and characterized two novel Mis18-interacting proteins, Eic1 and Eic2, in fission yeast. Eic1 is a small protein of approximately 13.4 kDa that is well conserved within *Schizosaccharomyces* species (figure 1*b*). The absence of an Eic2 orthologue in *S. japonicus* suggests that the primary sequence of Eic2 may have rapidly diverged over evolutionary time so that it is now undetectable within the *S. japonicus* genome; alternatively, other proteins may undertake the function of Eic2 in *S. japonicus*. The lack of any specific domains within Eic1 and Eic2 hinders the detection of similar or related proteins in other organisms, and thus orthologues of Eic1 and Eic2 remain unidentified outside of the *Schizosaccharomyces* clade.

The human Mis18 complex (consisting of hMis18α/β and Mis18BP1^{KNL2}) primes CENP-A assembly by associating with centromeres specifically in telophase, subsequently recruiting the CENP-A-specific chaperone HJURP^{Scm3} to mediate CENP-A deposition in G1 [13,25,26,28,29]. The phosphorylation of Mis18BP1^{KNL2} by CDK1/2 has been shown to regulate the recruitment of Mis18BP1^{KNL2} and timing of CENP-A assembly during the cell cycle [39]. In fission yeast, the known Cnp1^{CENP-A} loading factors, Mis16$^{RbAp46/48/Hat2}$, Mis18 and Scm3^{HJURP}, also exhibit a very specific cell-cycle-regulated localization to centromeres in that they dissociate before metaphase and re-associate in mid-anaphase [9,11–13]. Our analyses demonstrate that both Eic1 and Eic2 are centromere-specific proteins with cell-cycle dynamics that are very similar to that of Mis16$^{RbAp46/48/Hat2}$, Mis18 and Scm3^{HJURP} (figure 2 and electronic supplementary material, figure S2). The re-association of these five factors in

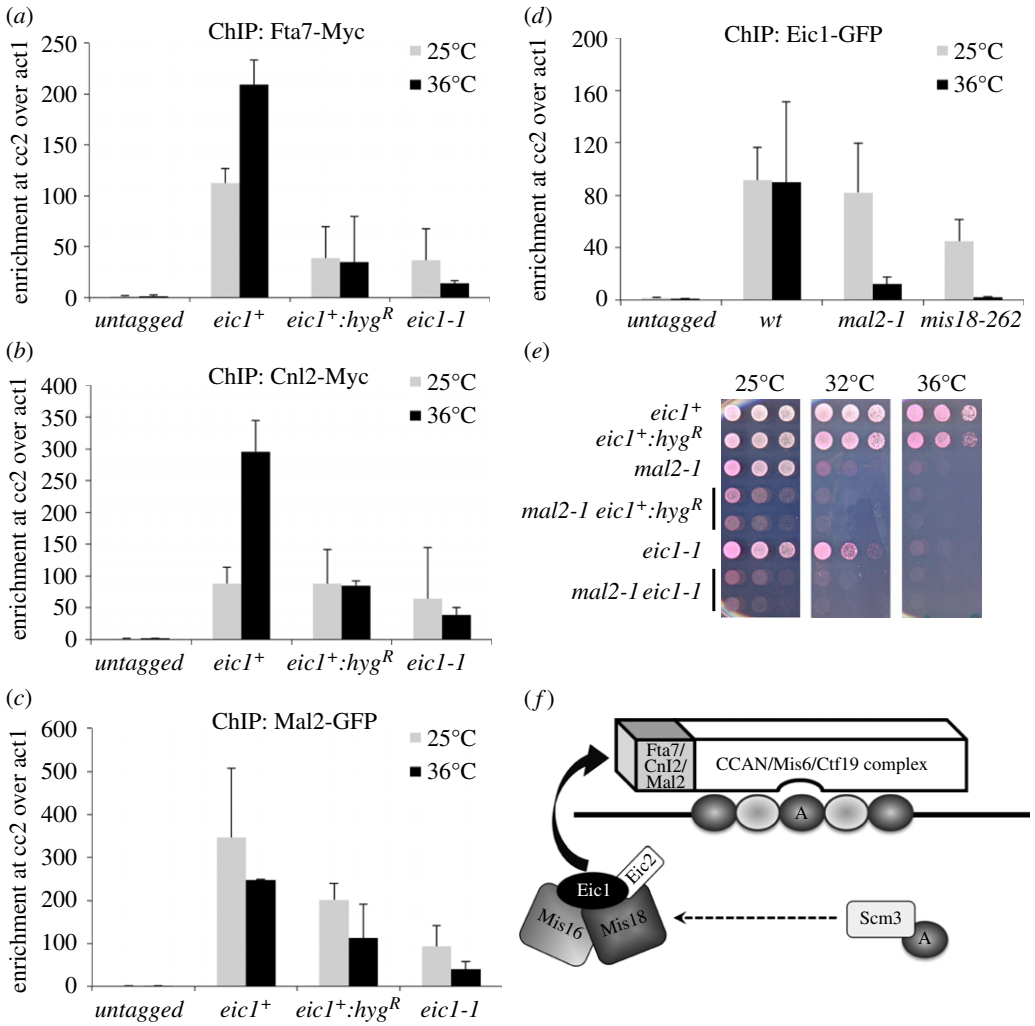


Figure 8. The CCAN components Fta7$^{\text{CENP-Q/Okp1}}$, Cnl2$^{\text{Nkp2}}$ and Mal2$^{\text{CENP-O/Mcm21}}$ influence Eic1 association with centromeres. (*a*) Fta7-Myc, (*b*) Cnl2-Myc and (*c*) Mal2-GFP association with centromeres is partly dependent on Eic1. (*d*) Eic1-GFP association with centromeres is greatly dependent on Mal2. qChIP analyses of (*a*) Fta7-Myc, (*b*) Cnl2-Myc, (*c*) Mal2-GFP and (*d*) Eic1-GFP association with centromeres in the indicated strains when grown at permissive (25°C) versus restrictive temperature (36°C) for 8 h. Enrichment of cc2 DNA relative to the *act1* locus is presented. Error bars represent standard deviation between at least three biological replicates. (*e*) *eic1-1* displays a severe negative genetic interaction when combined with *mal2-1*. Five-fold serial dilutions of cells spotted on YES + Phloxine B media and incubated at the indicated temperatures; dead cells stain dark pink. (*f*) A model for Cnp1$^{\text{CENP-A}}$ maintenance at *S. pombe* centromeres mediated by Eic1. The CCAN components Fta7$^{\text{CENP-Q/Okp1}}$, Cnl2$^{\text{Nkp2}}$ and Mal2$^{\text{CENP-O/Mcm21}}$ that are constitutively bound to centromeres together form a module that recruits Eic1, and thereby regulates the temporal association of Eic1, Mis16$^{\text{RbAp46/48/Hat2}}$, Mis18 and Eic2 with centromeres. Once bound, Mis16$^{\text{RbAp46/48/Hat2}}$/Mis18 then likely recruit the Cnp1$^{\text{CENP-A}}$-specific chaperone Scm3$^{\text{HJURP}}$ to centromeres (the dashed arrow indicates that only an *in vitro* association between these proteins has been demonstrated), and thus ensures replenishment of Cnp1$^{\text{CENP-A}}$ ('A' in closed oval) at centromeres in every cell cycle.

anaphase–telophase is presumably required to allow the subsequent replication-independent deposition of Cnp1$^{\text{CENP-A}}$ in G2 phase of the cell cycle [40,41]. Recent analysis of *Arabidopsis* Mis18BP1$^{\text{KNL2}}$ suggests that it conforms to the *S. pombe*, rather than the vertebrate, pattern of CENP-A loading factor recruitment to centromeres during the cell cycle [33].

Our analyses show that functional Eic1 is required for the association of its interacting partners Mis16$^{\text{RbAp46/48/Hat2}}$ and Mis18, as well as Scm3$^{\text{HJURP}}$ with centromeres (figure 4*a–c*). The *eic1*$^{+}$*:hyg*^{R} and *eic1-1* alleles also display strong negative interactions when combined with mutant Cnp1$^{\text{CENP-A}}$ or mutant Cnp1$^{\text{CENP-A}}$ assembly factors, Mis16$^{\text{RbAp46/48/Hat2}}$, Mis18 and Scm3$^{\text{HJURP}}$ (figure 5*a*; summarized in figure 5*c*). Such interactions are consistent with Eic1 being required to promote Cnp1$^{\text{CENP-A}}$ assembly by mediating the association of the Cnp1$^{\text{CENP-A}}$ chaperone Scm3$^{\text{HJURP}}$ as well as Mis16$^{\text{RbAp46/48/Hat2}}$ and Mis18 with centromeres. Eic1 is required to recruit Mis18 to centromeres, and functional Mis18 has been shown to affect the localization of the CCAN/Mis6/Ctf19 complex component Mis6$^{\text{CENP-I/Ctf3}}$ to centromeres [11]. It was thus surprising that our analyses only detected a partial reduction in Mis6$^{\text{CENP-I/Ctf3}}$ levels at centromeres in *eic1* mutant cells (figure 4*d*). Nevertheless, consistent with this finding, our analyses show that the association of other CCAN components with centromeres is only partly dependent on Eic1 function (Fta7$^{\text{CENP-Q/Okp1}}$, Cnl2$^{\text{Nkp2}}$ and Mal2$^{\text{CENP-O/Mcm21}}$; figure 8*a–c*). The CCAN proteins Mis6$^{\text{CENP-I/Ctf3}}$, Sim4$^{\text{CENP-K/Mcm22}}$, Mis15$^{\text{CENP-N/Chl4}}$ and Mis17$^{\text{CENP-U/Ame1}}$ have been shown to facilitate Cnp1$^{\text{CENP-A}}$ assembly [8,10,11]. Thus, although functional Eic1 is essential for the recruitment of the key Cnp1$^{\text{CENP-A}}$ assembly factors Mis16$^{\text{RbAp46/48/Hat2}}$, Mis18 and Scm3$^{\text{HJURP}}$ to centromeres, it is partly dispensable for maintaining constitutive CCAN components at centromeres.

Eic2, the second Mis18-interacting protein that we identified, contributes to Mis18 function (figure 5*b*) as it promotes normal levels of Mis18, Mis16$^{\text{RbAp46/48/Hat2}}$ and Mis6$^{\text{CENP-I/Ctf3}}$ association with centromeres (figure 4*e,f,h*). However, Eic2 is

dispensable for maintaining Cnp1^{CENP-A} and Scm3^{HJURP} association with centromeres (figures 3*d* and 4*g*). It was therefore unexpected that loss of Eic2 (*eic2*Δ) resulted in the partial rescue of *scm3-139* temperature sensitivity (figure 5*b*; summarized in figure 5*c*). Given that the levels of Scm3^{HJURP} protein and its association with centromeres remain unaffected in *eic2*Δ cells (figure 4*g*; electronic supplementary material, figure S6*c*), we speculate that Scm3^{HJURP}–Mis18 interactions are perhaps stabilized in the absence of Eic2. Regardless of these changes, the levels of loading factors remaining at centromeres in *eic2*Δ cells are sufficient to maintain normal Cnp1^{CENP-A} levels. The fact that the localization and function of Mis18, but not Eic1 or Cnp1^{CENP-A} (figures 3*d*, 4*e*, 5*b* and 6*b*), is compromised in *eic2*Δ cells suggests that Eic1 and Eic2 function independently. Eic2 may act to bolster Mis18 function under particular conditions of stress.

Eic1 association with centromeres is entirely dependent on functional Mis18, but only partly dependent on functional Mis16$^{RbAp46/48/Hat2}$ and Scm3^{HJURP}. Moreover, Eic1 levels at centromeres remain largely unaffected when Cnp1^{CENP-A}, Mis6$^{CENP-I/Ctf3}$ or Mis12 function is compromised (figure 6*a* and electronic supplementary material, figure S7*a*). These and other observations (figure 4*a–d*) suggest that Eic1 acts in concert with Mis16$^{RbAp46/48/Hat2}$ and Mis18 to promote recruitment of the Cnp1^{CENP-A}-specific chaperone Scm3^{HJURP} to centromeres and thereby facilitates Cnp1^{CENP-A} assembly. Thus, although no obvious protein similarity is evident between Eic1 and Mis18BP1^{KNL2}, we propose that Eic1 serves as the functional counterpart of Mis18BP1^{KNL2} [13, 33,37].

The maintenance of epigenetically specified centromeres on chromosomes requires a feedback mechanism where, once established, constitutive centromere proteins themselves recruit the CENP-A assembly factors and thereby ensure their preservation. In vertebrates, the CENP-A assembly factors HJURP^{Scm3}/Mis18/Mis18BP1^{KNL2} are recruited by the interaction of Mis18BP1^{KNL2} with CENP-C$^{Cnp3/Mif2}$ [27,28]. Our analyses suggest that Eic1 performs the equivalent function to Mis18BP1^{KNL2} by association with the Fta7$^{CENP-Q/Okp1}$, Cnl2^{Nkp2} and Mal2$^{CENP-O/Mcm21}$ subunits of the constitutive CCAN/Mis6/Ctf19 complex (figure 7). In support of this view, we find that Eic1 recruitment to centromeres is particularly dependent on functional Mal2$^{CENP-O/Mcm21}$, and *eic1 mal2* double mutants are severely compromised (figure 8*d*,*e*). By contrast, Eic1 association with centromeres is only partly dependent on the CCAN component Mis6$^{CENP-I/Ctf3}$ (figure 6*a* and electronic supplementary material, figure S7*a*). Mcm21$^{Mal2/CENP-O}$ and Okp1$^{Fta7/CENP-Q}$ along with Ctf19$^{Fta2/CENP-P}$ and Ame1$^{Mis17/CENP-U}$ are known to form COMA, a biochemically distinct subcomplex within the larger Ctf19 complex in *S. cerevisiae* [19,42], much like the stable CENP-O/P/Q/U subcomplex described in vertebrate cells [43]. We propose that Mal2$^{CENP-O/Mcm21}$, Fta7$^{CENP-Q/Okp1}$ and Cnl2^{Nkp2} form an analogous module within the CCAN/Mis6/Ctf19 complex that ensures the propagation of Cnp1^{CENP-A} chromatin and kinetochores by recruiting Eic1 and consequently the Cnp1^{CENP-A} assembly factors Mis16$^{RbAp46/48/Hat2}$, Mis18 and Scm3^{HJURP} to fission yeast centromeres.

Undoubtedly, the analyses of centromere–kinetochore architecture in distinct model systems have provided insights into the function of particular components. The lack of specific components in one organism reveals its reliance on alternative pathways. For example, *Drosophila* lack the kinetochore protein CENP-T$^{Cnp20/Cnn1}$, and consequently their kinetochores are completely reliant on the KMN (Knl1–Mis12–Ndc80) pathway for attachment to microtubules [4,18]. In vertebrates, Mis18BP1^{KNL2} directly associates with CENP-C$^{Cnp3/Mif2}$, allowing the maintenance and propagation of centromeric chromatin through the recruitment of Mis18, HJURP^{Scm3} and thus CENP-A [27,28]. Remarkably, the CENP-A^{CID} assembly pathway appears to be completely rewired during evolution of *Drosophila* because the main CENP-A loading factors HJURP^{Scm3}, Mis18 and Mis18BP1^{KNL2} have been lost and replaced by the Cal1 protein, which directly associates with CENP-C$^{Cnp3/Mif2}$ [31,32]. In *S. cerevisiae*, where centromere specification is driven by the recognition of specific DNA elements by DNA-binding proteins [15], a self-propagation mechanism is no longer necessary and consequently proteins equivalent to Mis18/Mis18BP1^{KNL2}, Cal1 or Eic1 are absent. Interestingly, fission yeast Cnp3$^{CENP-C/Mif2}$ is not essential indicating that it is not required for the propagation of Cnp1^{CENP-A} chromatin. Eic1 appears to perform an equivalent role to Mis18BP1^{KNL2} but associates with the essential constitutive CCAN components Mal2$^{CENP-O/Mcm21}$, Fta7$^{CENP-Q/Okp1}$ and Cnl2^{Nkp2}, rather than Cnp3$^{CENP-C/Mif2}$. Thus, it is possible that the highly conserved CCAN components CENP-O$^{Mal2/Mcm21}$ and CENP-Q$^{Fta7/Okp1}$ also function in other organisms to recruit CENP-A assembly factors; they may even have an equivalent underlying function that is redundant with CENP-C$^{Cnp3/Mif2}$ at vertebrate centromeres.

In conclusion, our analysis has identified Eic1 as a factor that connects Mis16$^{RbAp46/48/Hat2}$ and Mis18, which are required for Cnp1^{CENP-A} incorporation, with constitutive kinetochore components within the CCAN/Mis6/Ctf19 complex. We propose that Eic1 serves two major functions: (i) priming Cnp1^{CENP-A} assembly in concert with Mis16$^{RbAp46/48/Hat2}$ and Mis18 and (ii) promoting kinetochore integrity in conjunction with the CCAN/Mis6/Ctf19 complex. The interdependency relationships among Eic1, Mis16$^{RbAp46/48/Hat2}$ and Mis18, along with their similar dynamic localization to centromeres, indicate that Eic1, Mis16$^{RbAp46/48/Hat2}$ and Mis18 form a complex that is temporarily released from centromeres during mitosis and can associate with centromeres independently of Cnp1^{CENP-A}. The novel physical interaction via Eic1 that we have uncovered between constitutively bound CCAN components and Mis18 is likely to be fundamental for the temporal regulation of Eic1–Mis16$^{RbAp46/48/Hat2}$–Mis18 association with centromeres, and thereby the cell-cycle-dependent assembly of Cnp1^{CENP-A} chromatin. The identification of Eic1 and Eic2 adds to the repertoire of known cell-cycle-regulated centromeric proteins involved in maintaining CENP-A at centromeres and provides a good example of how distinctly different proteins can contribute to a conserved cellular process in diverse organisms.

5. Material and methods

5.1. Yeast strains and standard techniques

Standard methods were used for fission yeast growth, genetics and manipulation [44]. Five-fold serial dilutions of the indicated strains were spotted onto YES media containing Phloxine B for growth assays, or DMSO or 12.5 µg ml^{-1} TBZ for TBZ sensitivity assays. Gene deletions, tagging and centromeric plasmid transformations were carried out by either the lithium acetate transformation method or electroporation. The *eic1*$^{+}$*:hyg*^{R} allele was generated by integrating a hygromycin resistance marker within the 3′UTR of *eic1*$^{+}$

at its endogenous locus. The *eic1-1* and *eic1-2* alleles were generated likewise, but derived by error-prone PCR using the GeneMorph II random mutagenesis kit (Agilent Technologies). The epitope-tagged alleles of Eic1, Eic2, Hat1, Fta7 and Cnl2 were generated at their respective endogenous loci by integrating an in-frame GFP, 3HA or 13myc cassette at their respective C termini [45].

5.2. Immunoaffinity purification and mass spectrometry

For Mis16 and Mis18 pulldowns, 5 g of pulverized *S. pombe* cells expressing Myc-tagged Mis16 or Mis18 were used for immunoprecipitation with anti-Myc antibody 9E10 (Covance) coupled to Protein G Dynabeads (Life Technologies), alongside an untagged control. For Eic1 and Eic2 pulldowns, 5 g of pulverized *S. pombe* cells expressing GFP-tagged Eic1 or Eic2 were used for immunoprecipitation with anti-GFP antibody A11122 (Life Technologies) coupled to Protein G Dynabeads, alongside an untagged control. After washes, Dynabeads with immunoprecipitated material were subjected to on-bead tryptic digestion, following which the samples were treated as described [9]. The average number of unique peptides corresponding to proteins that were reproducibly enriched in the epitope-tagged samples but consistently absent in the untagged controls, over three independent biological replicates, is presented.

5.3. Multiple sequence alignment

Orthologues of Eic1 and Eic2 among the *Schizosaccharomyces* species were identified by BLAST, PSI-BLAST or synteny searches against the *Schizosaccharomyces* group database available at the Broad Institute *Schizosaccharomyces* Comparative Genome Project [46]. Primary sequences of Eic1 or Eic2 orthologues were aligned using ClustalW and T-Coffee.

5.4. Co-immunoprecipitations and western analyses

For co-immunoprecipitation experiments, 2 g of pulverized *S. pombe* cells expressing the indicated epitope-tagged proteins were used for immunoprecipitation using anti-GFP antibody A11122 (Life Technologies), anti-Myc antibody 9E10 (Covance), anti-Myc antibody 9B11 (Cell Signaling) or anti-HA antibody 12CA5 (Roche) coupled to Protein G Dynabeads (Life Technologies). The same antibodies were also used for western analyses as indicated. Anti-Bip1 was used as a loading control where indicated.

5.5. Recombinant protein co-expression and binding assays

Codon-optimized ORFs of Eic1, Mis16 and Mis18 were synthesized by GeneArt (Life Technologies), for expression in *E. coli*. These were then sub-cloned into pGEX-6P-1 (GST) (GE Healthcare, gift from J. Welburn) or pEC(K)-3CHis (His) (gift from A. A. Jeyaprakash) expression vectors as indicated, and co-transformed into the BL21-Gold-pLysS *E. coli* strain (Agilent Technologies). Co-transformants were grown in SuperBroth supplemented with carbenicillin and kanamycin as appropriate, and the indicated proteins co-expressed by induction with 0.3 mM IPTG. GST pulldowns were done using glutathione agarose (Sigma). His pulldowns were done using Ni-NTA agarose beads (Qiagen). Samples were resolved on NuPAGE Bis-Tris gels (Life Technologies) and stained using InstantBlue (Expedeon).

5.6. Quantitative chromatin immunoprecipitation

The indicated *S. pombe* strains were grown in YES media at 32°C. If appropriate, cells were shifted to restrictive temperature (36°C) for 8 h, or continued to grow for the same length of time at permissive temperature (25°C) before fixation. For ChIPs on centromeric plasmids, cells harbouring pH-cc2 [38,47] were grown in PMG media minus adenine, minus uracil, at 32°C. To confirm that plasmids were behaving episomally and had not integrated, a plasmid stability test was performed at the time of fixation. Cells (100–1000) were plated onto YES supplemented with 1/10th adenine and allowed to form colonies. Samples exhibiting no integrations were used for ChIP.

ChIP was performed as described [9], or with the following modifications. Cells were fixed with 1% formaldehyde (Sigma) for 20 min at room temperature and lysed using a bead beater (Biospec Products). Cell lysates were sonicated in a Bioruptor (Diagenode) (15 min, 30 s On and 30 s Off at 'High' (200 W) position). For $Cnp1^{CENP\text{-}A}$ ChIPs, anti-$Cnp1^{CENP\text{-}A}$ antiserum was used with Protein G agarose beads (Roche). For all other ChIPs, Protein G Dynabeads (Life Technologies) were used along with anti-GFP A11122 (Life Technologies), anti-HA 12CA5 (Roche) or anti-Myc 9B11 (Cell Signaling), as appropriate. Immunoprecipitated DNA was recovered using the Chelex-100 resin (BioRad) [48]. ChIPs were analysed by real-time PCR using Lightcycler 480 SYBR Green (Roche) with primers specific to the central cores of centromere 2 (*cc2*) or centromeres 1/3 (*cc1/3*), the innermost repeats of centromere 1 (*imr*) or *act1*. All ChIP enrichments were calculated as % DNA immunoprecipitated at the locus of interest relative to the corresponding input samples, and normalized to % DNA immunoprecipitated at the *act1* locus. Histograms represent data averaged over at least three biological replicates. Error bars represent standard deviations from at least three biological replicates.

5.7. ChIP-seq analysis

Schizosaccharomyces pombe strains expressing the indicated GFP-tagged proteins were grown in YES media at 32°C to 1.25×10^9 cells at a density of 1×10^7 cells ml^{-1}. Cells were fixed for 15 min in 1% formaldehyde (Sigma) and lysed with 0.4 mg ml^{-1} Zymolyase 100 T (AMS Biotechnology Europe) in PEMS for 1 h at 36°C. Cell lysates were sonicated in a Bioruptor (Diagenode) (20 min, 30 s On and 30 s Off at 'High' (200 W) position) and immunoprecipitated overnight using anti-GFP antibody A11122 (Life Technologies) and Protein G Dynabeads (Life Technologies). The samples were washed and cross-links reversed using 1% SDS for 4 h at 65°C. Immunoprecipitated DNA was recovered using a Qiagen PCR purification kit, and centromeric enrichment was verified by qPCR using Lightcycler 480 SYBR Green (Roche). Illumina libraries were prepared following the TruSeq Nano DNA kit (Illumina) guidelines using NEXTflex (Bio Scientific) adapters with internal barcodes. Multiplexed libraries were 100 bp paired-end sequenced on an Illumina HiSeq2000 (Ark Genomics, Edinburgh, UK). ChIP-seq data were mapped onto the *S. pombe* genome assembly EF2

(Ensemble) using Bowtie2. Coverage calculations and peak calling were done using MACS and PeakSplitter.

5.8. Cytology

Immunolocalization and microscopy were performed as described [9]. If appropriate, cells were shifted to restrictive temperature (36°C) for 8 h before fixation. Cells were fixed for 7–10 min with 3.7% formaldehyde (Sigma). Fixation of cells for tubulin staining used formaldehyde and 0.06% glutaraldehyde. Antibodies used were anti-GFP A11122 (1 : 200) (Life Technologies), TAT1 anti-tubulin (1 : 15) (gift from K. Gull) or anti-Cnp1^{CENP-A} antiserum (1 : 1000). Alexa Fluor 594- and 488-coupled secondary antibodies were used at 1 : 1000 (Life Technologies).

5.9. Centromeric plasmid selection system and stability

pH-cc2 (H denotes an *otr* heterochromatic element and cc denotes central domain DNA) carries *ura4*$^+$ and *sup3-5* (suppressor of *ade6-704*) selection systems [38,47]. Cells without *ura4*$^+$ cannot grow on minus-uracil plates, while *ade6-704* cells do not grow without adenine and form red colonies on 1/10th adenine plates. The *sup3-5*-tRNA gene suppresses a premature stop in *ade6-704*, allowing growth on minus-adenine plates. Cells containing pH-cc2 form a high percentage of white or sectored colonies on 1/10th adenine indicator plates, demonstrating their relative mitotic stability in wild-type cells. In cells lacking Clr4, however, their mitotic stability is lost due to a lack of heterochromatin-dependent centromeric cohesion. To confirm that plasmids were behaving episomally and had not integrated, a plasmid stability test was performed at the time of fixation for ChIP. Cells (100–1000) were plated onto YES supplemented with 1/10th adenine and allowed to form colonies. Samples exhibiting no integrations were used for ChIP.

Note added in proof

A concurrent study has identified Eic1 and Eic2 as Mis19 and Mis20, respectively, in association with Mis18 [49].

Acknowledgements. We are grateful to F. de Lima Alves for mass spectrometry support, and A. A. Jeyaprakash for advice on recombinant protein co-expression and pulldowns. We also thank L. Sanchez-Pulido for attempting to bioinformatically find Eic1 and Eic2 orthologues. Additionally, we thank M. Yanagida and U. Fleig for strains; A. A. Jeyaprakash and J. Welburn for *E. coli* expression vectors; and A. Pidoux, E. Choi, S. Catania and members of the Allshire lab for advice, discussion and reagents. We also thank A. Marston and A. Pidoux for critical comments on the manuscript. L.S. performed experiments. N.R.T.T. performed ChIP-seq analyses. J.R. provided resources for mass spectrometry. L.S. and R.C.A. jointly conceived and designed the study, and wrote the manuscript.

Data accessibility. ChIP-seq data have been submitted to the Gene Expression Omnibus under accession no. GSE 54685.

Funding statement. L.S. was supported by an EC FP7 Marie Curie International Incoming Fellowship (PIIF-GA-2010-275280) and an EMBO Long Term Fellowship (ALTF 1491-2010). The Darwin Trust and a Principal's Career Development scholarship supported N.R.T.T. The Wellcome Trust supported the work of R.C.A. (095021 and 065061) and J.R. (084229) along with funding from the European Commission Network of Excellence EpiGeneSys (HEALTH-F4-2010-257082) to R.C.A. The Wellcome Trust Centre for Cell Biology (092076) and mass spectrometry instrumentation (091020) are supported by funding from the Wellcome Trust. R.C.A. is a Wellcome Trust Principal Research Fellow.

References

1. Scott KC, Sullivan BA. 2013 Neocentromeres: a place for everything and everything in its place. *Trends Genet.* **30**, 66–74. (doi:10.1016/j.tig.2013.11.003)
2. Stimpson KM, Matheny JE, Sullivan BA. 2012 Dicentric chromosomes: unique models to study centromere function and inactivation. *Chromosome Res.* **20**, 595–605. (doi:10.1007/s10577-012-9302-3)
3. Gomez-Rodriguez M, Jansen LE. 2013 Basic properties of epigenetic systems: lessons from the centromere. *Curr. Opin. Genet. Dev.* **23**, 219–227. (doi:10.1016/j.gde.2012.11.002)
4. Catania S, Allshire RC. 2014 Anarchic centromeres: deciphering order from apparent chaos. *Curr. Opin. Cell Biol.* **26**, 41–50. (doi:10.1016/j.ceb.2013.09.004)
5. Muller S, Almouzni G. 2013 A network of players in H3 histone variant deposition and maintenance at centromeres. *Biochim. Biophys. Acta* **1839**, 241–250. (doi:10.1016/j.bbagrm.2013.11.008)
6. Mendiburo MJ, Padeken J, Fulop S, Schepers A, Heun P. 2011 *Drosophila* CENH3 is sufficient for centromere formation. *Science* **334**, 686–690. (doi:10.1126/science.1206880)
7. Fachinetti D *et al*. 2013 A two-step mechanism for epigenetic specification of centromere identity and function. *Nat. Cell Biol.* **15**, 1056–1066. (doi:10.1038/ncb2805)
8. Takahashi K, Chen ES, Yanagida M. 2000 Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. *Science* **288**, 2215–2219. (doi:10.1126/science.288.5474.2215)
9. Pidoux AL *et al*. 2009 Fission yeast Scm3: a CENP-A receptor required for integrity of subkinetochore chromatin. *Mol. Cell* **33**, 299–311. (doi:10.1016/j.molcel.2009.01.019)
10. Pidoux AL, Richardson W, Allshire RC. 2003 Sim4: a novel fission yeast kinetochore protein required for centromeric silencing and chromosome segregation. *J. Cell Biol.* **161**, 295–307. (doi:10.1083/jcb.200212110)
11. Hayashi T, Fujita Y, Iwasaki O, Adachi Y, Takahashi K, Yanagida M. 2004 Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. *Cell* **118**, 715–729. (doi:10.1016/j.cell.2004.09.002)
12. Williams JS, Hayashi T, Yanagida M, Russell P. 2009 Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. *Mol. Cell* **33**, 287–298. (doi:10.1016/j.molcel.2009.01.017)
13. Fujita Y, Hayashi T, Kiyomitsu T, Toyoda Y, Kokubu A, Obuse C, Yanagida M. 2007 Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1. *Dev. Cell* **12**, 17–30. (doi:10.1016/j.devcel.2006.11.002)
14. Loyola A, Almouzni G. 2004 Histone chaperones, a supporting role in the limelight. *Biochim. Biophys. Acta* **1677**, 3–11. (doi:10.1016/j.bbaexp.2003.09.012)
15. Allshire RC, Karpen GH. 2008 Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat. Rev. Genet.* **9**, 923–937. (doi:10.1038/nrg2466)
16. Liu X, McLeod I, Anderson S, Yates III JR, He X. 2005 Molecular analysis of kinetochore architecture in fission yeast. *EMBO J.* **24**, 2919–2930. (doi:10.1038/sj.emboj.7600762)
17. Shiroiwa Y, Hayashi T, Fujita Y, Villar-Briones A, Ikai N, Takeda K, Ebe M, Yanagida M. 2011 Mis17 is a regulatory module of the Mis6-Mal2-Sim4 centromere complex that is required for the recruitment of CenH3/CENP-A in fission yeast. *PLoS ONE* **6**, e17761. (doi:10.1371/journal.pone.0017761)
18. Schleiffer A, Maier M, Litos G, Lampert F, Hornung P, Mechtler K, Westermann S. 2012 CENP-T proteins are conserved centromere receptors of the Ndc80 complex. *Nat. Cell Biol.* **14**, 604–613. (doi:10.1038/ncb2493)

19. Cheeseman IM, Anderson S, Jwa M, Green EM, Kang J, Yates III JR, Chan CS, Drubin DG, Barnes G. 2002 Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p. *Cell* **111**, 163–172. (doi:10.1016/S0092-8674(02)00973-X)
20. Foltz DR, Jansen LET, Black BE, Bailey AO, Yates JR, Cleveland DW. 2006 The human CENP-A centromeric nucleosome-associated complex. *Nat. Cell Biol.* **8**, 458–469. (doi:10.1038/ncb1397)
21. Santaguida S, Musacchio A. 2009 The life and miracles of kinetochores. *EMBO J.* **28**, 2511–2531. (doi:10.1038/emboj.2009.173)
22. Sanchez-Pulido L, Pidoux AL, Ponting CP, Allshire RC. 2009 Common ancestry of the CENP-A chaperones Scm3 and HJURP. *Cell* **137**, 1173–1174. (doi:10.1016/j.cell.2009.06.010)
23. Jansen LET, Black BE, Foltz DR, Cleveland DW. 2007 Propagation of centromeric chromatin requires exit from mitosis. *J. Cell Biol.* **176**, 795–805. (doi:10.1083/jcb.200701066)
24. Dunleavy EM, Almouzni G, Karpen GH. 2011 H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G(1) phase. *Nucleus* **2**, 146–157. (doi:10.4161/nucl.2.2.15211)
25. Foltz DR, Jansen LET, Bailey AO, Yates JR, Bassett EA, Wood S, Black BE, Cleveland DW. 2009 Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell* **137**, 472–484. (doi:10.1016/j.cell.2009.02.039)
26. Dunleavy EM, Roche D, Tagami H, Lacoste N, Ray-Gallet D, Nakamura Y, Daigo Y, Nakatani Y, Almouzni-Pettinotti G. 2009 HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* **137**, 485–497. (doi:10.1016/j.cell.2009.02.040)
27. Dambacher S *et al*.. 2012 CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. *Nucleus* **3**, 101–110. (doi:10.4161/nucl.18955)
28. Moree B, Meyer CB, Fuller CJ, Straight AF. 2011 CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. *J. Cell Biol.* **194**, 855–871. (doi:10.1083/jcb.201106079)
29. Barnhart MC, Kuich PH, Stellfox ME, Ward JA, Bassett EA, Black BE, Foltz DR. 2011 HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J. Cell Biol.* **194**, 229–243. (doi:10.1083/jcb.201012017)
30. Tanaka K, Chang HL, Kagami A, Watanabe Y. 2009 CENP-C functions as a scaffold for effectors with essential kinetochore functions in mitosis and meiosis. *Dev. Cell* **17**, 334–343. (doi:10.1016/j.devcel.2009.08.004)
31. Phansalkar R, Lapierre P, Mellone BG. 2012 Evolutionary insights into the role of the essential centromere protein CAL1 in *Drosophila*. *Chromosome Res.* **20**, 493–504. (doi:10.1007/s10577-012-9299-7)
32. Chen CC, Dechassa ML, Bettini E, Ledoux MB, Belisario C, Heun P, Luger K, Mellone BG. 2014 CAL1 is the *Drosophila* CENP-A assembly factor. *J. Cell Biol.* **204**, 313–329. (doi:10.1083/jcb.201305036)
33. Lermontova I, Kuhlmann M, Friedel S, Rutten T, Heckmann S, Sandmann M, Demidov D, Schubert V, Schubert I. 2013 *Arabidopsis* kinetochore null2 is an upstream component for centromeric histone H3 variant cenH3 deposition at centromeres. *Plant Cell* **25**, 3389–3404. (doi:10.1105/tpc.113.114736)
34. Lermontova I, Rutten T, Schubert I. 2011 Deposition, turnover, and release of CENH3 at *Arabidopsis* centromeres. *Chromosoma* **120**, 633–640. (doi:10.1007/s00412-011-0338-5)
35. Tong K, Keller T, Hoffman CS, Annunziato AT. 2012 *Schizosaccharomyces pombe* Hat1 (Kat1) is associated with Mis16 and is required for telomeric silencing. *Eukaryot. Cell* **11**, 1095–1103. (doi:10.1128/EC.00123-12)
36. Parthun MR. 2007 Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene* **26**, 5319–5328. (doi:10.1038/sj.onc.1210602)
37. Maddox PS, Hyndman F, Monen J, Oegema K, Desai A. 2007 Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. *J. Cell Biol.* **176**, 757–763. (doi:10.1083/jcb.200701065)
38. Folco HD, Pidoux AL, Urano T, Allshire RC. 2008 Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. *Science* **319**, 94–97. (doi:10.1126/science.1150944)
39. Silva MC, Bodor DL, Stellfox ME, Martins NM, Hochegger H, Foltz DR, Jansen LE. 2012 Cdk activity couples epigenetic centromere inheritance to cell cycle progression. *Dev. Cell* **22**, 52–63. (doi:10.1016/j.devcel.2011.10.014)
40. Lando D *et al*.. 2012 Quantitative single-molecule microscopy reveals that CENP-A^{Cnp1} deposition occurs during G2 in fission yeast. *Open Biol.* **2**, 120078. (doi:10.1098/rsob.120078)
41. Takayama Y, Sato H, Saitoh S, Ogiyama Y, Masuda F, Takahashi K. 2008 Biphasic incorporation of centromeric histone CENP-A in fission yeast. *Mol. Biol. Cell* **19**, 682–690. (doi:10.1091/mbc.E07-05-0504)
42. De Wulf P, McAinsh AD, Sorger PK. 2003 Hierarchical assembly of the budding yeast kinetochore from multiple subcomplexes. *Genes Dev.* **17**, 2902–2921. (doi:10.1101/gad.1144403)
43. Hori T, Okada M, Maenaka K, Fukagawa T. 2008 CENP-O class proteins form a stable complex and are required for proper kinetochore function. *Mol. Biol. Cell* **19**, 843–854. (doi:10.1091/mbc.E07-06-0556)
44. Moreno S, Klar A, Nurse P. 1991 Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823.
45. Bahler J, Wu JQ, Longtine MS, Shah NG, McKenzie III A, Steever AB, Wach A, Philippsen P, Pringle JR. 1998 Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**, 943–951. (doi:10.1002/(SICI)1097-0061(199807)14:10<943::AID-YEA292>3.0.CO;2-Y)
46. Rhind N *et al*. 2011 Comparative functional genomics of the fission yeasts. *Science* **332**, 930–936. (doi:10.1126/science.1203357)
47. Baum M, Ngan VK, Clarke L. 1994 The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharomyces pombe* centromere. *Mol. Biol. Cell* **5**, 747–761. (doi:10.1091/mbc.5.7.747)
48. Nelson JD, Denisenko O, Bomsztyk K. 2006 Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat. Protoc.* **1**, 179–185. (doi:10.1038/nprot.2006.27)
49. Hayashi T, Ebe M, Nagao K, Kokubu A, Sajiki K, Yanagida M. In press. *Schizosaccharomyces pombe* centromere protein Mis19 links Mis16 and Mis18 to recruit CENP-A through interacting with NMD factors and the SWI/SNF complex. *Genes Cells* (doi:10.1111/gtc.12152)